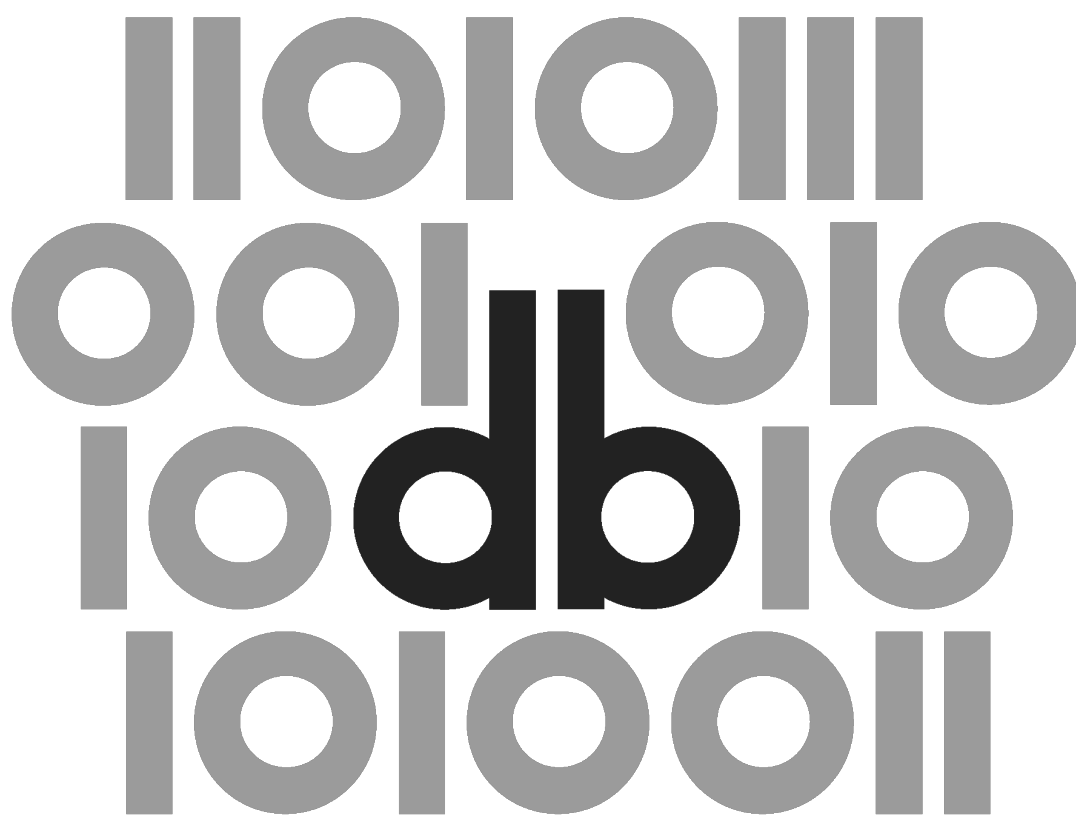


3 (2020)

<DIGITÁLIS BÖLCSÉSZET>

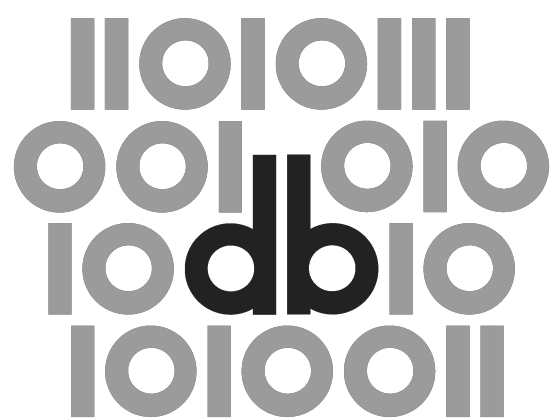


3 (2020)

</DIGITÁLIS BÖLCSÉSZET>

**Digitális Bölcsészet**  
**2020., harmadik szám**

# <DIGITÁLIS BÖLCSÉSZET>



3 (2020)

**Felelős szerkesztő:**

Maróthy Szilvia

**Szerkesztőség:**

Kokas Károly, Parádi Andrea

**Rovatvezetők:**

*Tanulmányok:* Kiss Margit

*Műhely:* Péter Róbert

*Kritika:* Almási Zsolt

*Labor:* Maróthy Szilvia

**Tanácsadó testület:**

Bartók István, Fazekas István, Golden Dániel, Horváth Iván, Palkó Gábor, Pap Balázs,  
Sass Bálint, Seláf Levente

**Korábbi munkatársaink:**

Bartók Zsófia Ágnes (szerkesztő, rovatvezető), Fodor János (szerkesztő), †Labádi  
Gergely (szerkesztő, rovatvezető), †Orlovsky Géza (tanácsadó testület)

**ISSN 2630-9696**

**DOI 10.31400/dh-hun.2020.3**

Kiadja a Bakonyi Géza Alapítvány és az ELTE BTK Régi Magyar  
Irodalom Tanszéke (1088 Budapest, Múzeum krt. 4/A).

Felelős kiadó az ELTE BTK Régi Magyar Irodalom Tanszék vezetője.

Megjelenik az Open Journal Systems (OJS) v. 3. platformon, melynek  
működtetését az ELTE Egyetemi Könyvtár- és Levéltár biztosítja.



Ez a mű a Creative Commons *Nevezd meg! – Ne add el! – Így add tovább! 2.5 Magyarországi Licenc* (<http://creativecommons.org/licenses/by-nc-sa/2.5/hu/>) feltételeinek megfelelően felhasználható.

Honlap: <http://ojs.elte.hu/digitalisbolcseszett>

Email cím: [dbfolyoirat@gmail.com](mailto:dbfolyoirat@gmail.com)

Olvasószerkesztő: Bucsecs Katalin

Tördelés: Hegedüs Béla

Grafika: Hegyi Gábor

# Tartalom

<b>TANULMÁNYOK</b>	<b>1</b>
Fogarasi György	
Eszkalálódott olvasás	
Stanford, Benjamin, Poe . . . . .	3
Fülöp Endre	
Szemantikus katalógus mint digitális bölcsészet . . . . .	21
Ring Orsolya–Kiss László	
Agrárpolitikai kihívások és jogszabályalkotás a korai Kádár-korban	
Történeti források szövegelemzése és szövegbányászati vizsgálata . . .	37
<b>MŰHELY</b>	<b>1</b>
Horváth Péter	
A vershangzás jellemzőinek automatikus feltárása József Attila verseiben	3
Harangozó Ádám	
FactGrid: adatbázis történészeknek . . . . .	29
<b>LABOR</b>	<b>1</b>
Kiss Tamás–Kanász-Nagy Zoltán	
Rekordoldali médiamegjelenítés a <i>DSpace</i> 7-ben . . . . .	3
Tamás Kiss–Zoltán Kanász-Nagy	
Record Page Media Player in <i>DSpace</i> 7 . . . . .	11
<b>KRITIKA</b>	<b>1</b>
Olivier Le Deuff. <i>Digital Humanities: History and Development</i> (2018) –	
Maróthy Szilvia . . . . .	3
Bolya Máttyás. <i>Információelmélet és népzene kutatás: Rendszeralkotás, nyilván-         vántartás, digitális archívum</i> (2019) – Király Péter . . . . .	7



<TANULMÁNYOK>





Fogarasi György

Szegedi Tudományegyetem

fogarasi@hung.u-szeged.hu

## Eszkálálódott olvasás Stanford, Benjamin, Poe

Az utóbbi évtizedek távoli *versus* szoros olvasásról folytatott vitái általában a léptékek körül forogtak. A vitában résztvevők vagy a távolság (avagy sebesség), vagy pedig a közelség (avagy lassúság) mellett szálltak síkra. Egy metakritikai szinten egyesek amellet érveltek, hogy minden olvasásnak képesnek kell lennie a különböző léptékek váltogatására vagy kombinálására. Kevés szó esett ugyanakkor az efféle skálázás korlátairól, és arról, hogy az olvasás mennyire nem redukálható a skála-logikára. A közeli és a távoli olvasás szószólóinak néhány elgondolkodtató megfogalmazásából kiindulva ez a tanulmány a skálázásban rejlő lehetőségeket és buktatókat próbálja feltárni. Először „Stanford” (mint egyetem és mint alapító) lesz fókuszban, aztán Walter Benjamin filmmel kapcsolatos fejtegetése kerül terítékre, végül a tanulmány Poe *Az ellopott levél* című detektívtörténetének pár részletét olvassa újra. Ez a három hivatkozási pont (Stanford, Benjamin, Poe) egymással párhuzamba állítható, amennyiben egyszerre hangsúlyozzák a szerializálást és a szegmentálást, a gyorsítást és a lassítást, avagy a távolságot és a közelséget. Közelebbről (vagy távolabbról?) nézve azonban Poe szövege túl is lép a skálázás eme logikáján. Olyan detekciós elvről tanúskodik, amely meghaladja a merőben be- vagy kizúmolásra építő stratégiákat, és az olvasás olyan fogalma felé mutat, amely nem csupán rendkívüli sebességbeli vagy távolságbeli léptéke miatt mondható „eszkálálódottnak”, hanem – elemibb szinten – azért, mert miközben továbbra is léptékekhez köti az olvasást, olyan vetülete is van, mely teljességgel kívül esik a skálázás bármiféle logikáján. A tanulmány az olvasásnak erre a radikális értelemben „eszkálálódott” (léptékeken túli) aspektusára igyekszik rámutatni.

Kulcsszavak:

szoros olvasás, távoli olvasás, keresés, lépték, skála, eskaláció



Az eskaláció fogalmát rendszerint traumatikus történelmi incidensek kontextusában használjuk, az erőszak váratlan kitöréseire, különösen, amikor az erőszakos cselekmények gyorsuló sorozatot alkotnak, egyfajta szélesedő spirált, amely azzal fenyeget, hogy magába szippantja nemcsak az adott konfliktus közvetlen résztvevőit, hanem a nézőket is, sőt veszélyezteti a közvetítés kereteit és a béke bevett intézményeit is. A mögöttes gondolat ilyenkor az, hogy a szélsőséges reaktivitás efféle eseteiben a cselekvés normális vagy szokványos szintjét rendkívüli léptékek váltják fel. Pontosan

erre utal maga a kifejezés: az események olyankor „e-szkalálódnak”, amikor a megszokott „léptéken kívülre” kerülnek, s ezzel kikerülnek a mindennapi irányítás alól. Ebből fakad az is, hogy a gyakori felszólítás a „de-eszkalációra” nem csupán józanságra vagy önuralomra int, hanem tárgyalásra és megbékélésre is szólít. Szorosabb, bár szokatlanabb értelemben véve azonban az „eszkaláció” kifejezés a léptékek (azaz bármely konkrét lépték) teljes irrelevanciájára is utalhat, egy olyan működés irányába mutatva, amely abban a radikális értelemben nevezhető eszkalálódottnak, hogy nemcsak ezt vagy azt a hétköznapi léptéket haladja meg, hanem magát a léptéket mint olyat. A jelen tanulmány ez utóbbi lehetőség kidolgozásához próbál hozzájárulni.

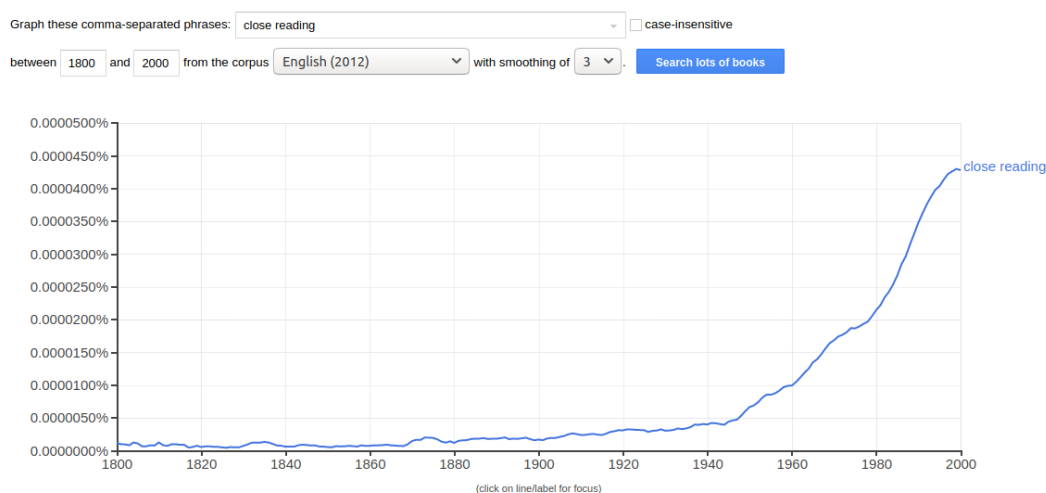
## 1. Szoros és távoli olvasás

Az archívumok számítógépes segédlettel végrehajtott tömeges elemzésével az irodalomértelmezés az utóbbi évtizedekben sosem látott szintekre lépett. Az új eljárások mind távolság, mind sebesség tekintetében messze meghaladták az olvasás szokványos léptékeit. Az olvasás úgyszólván eszkalálódott. E fejlemény nyomán a „távoli” kontra „szoros” olvasás körüli vita – az „olvasás körüli háborúskodás” (*reading wars*), ahogy egyesek nevezik<sup>1</sup> – a léptékek kérdése körül forgott. A vitázó felek legtöbbször vagy a távolság (vagy sebesség) védelmére keltek, vagy a közelséget (vagy lassúságot) pártolták. Olyanok is akadtak, akik – afféle metakritikai szinten – amellet érveltek, hogy minden olvasásnak képesnek kell lennie a skálázásra, az elemzés különböző léptékei közti váltásra, a léptékek kombinálására. Kevés szó esett viszont magának a skálázásnak a hatáiról, és arról, hogy az olvasás nem korlátozható a léptékek logikájára. Az alábbiakban a szoros és a távoli olvasás főbb szószólóinak néhány elgondolkodtató megfogalmazásából kiindulva a skálázásban rejlő lehetőségek és korlátok vizsgálatát célozom meg, először néhány utalással „Stanfordra” (az egyetemre és annak alapítójára), majd megvizsgálva Walter Benjamin filmre vonatkozó megjegyzéseit, végül pedig – ami talán a leglényegesebb – újraolvasva Poe *Az ellopott levél* című detektívtörténetének néhány részletét. Ez a három hivatkozási pont (Stanford, Benjamin és Poe) hasonlónak tűnik abból a szempontból, hogy mindhárman egyszerre hangsúlyozzák a szerializálás és a szegmentálás, a gyorsítás és a lassítás, vagy a távolság és a közelség fontosságát. Közelebről (vagy talán távolabbról?) nézve azonban Poe szövege tovább is megy, és túlmutat a skálázás iménti sémáján. A detektálás olyan logikájáról tanúskodik, amely túllép a bezúmolási és kizúmolási stratégiákon, és az olvasás egy olyan fogalma felé mutat, mely nem egyszerűen a sebesség vagy távolság terén elért rendkívüli léptéke miatt nevezhető eszkalálódottnak, hanem – radikálisabb értelemben – azért, mert teljességgel más logikát követ, mint a skálázás. Az olvasás ekként nem csupán abban az értelemben válhat „eszkalálódottá”, hogy túlhalad a *hétköznapi* léptéken, hanem abban az értelemben is, hogy – miközben továbbra is kötődik konkrét léptékekhez – bizonyos szempontból *minden* léptéket meghalad, sőt meghaladja a léptéket *mint olyat*.

A „skála retorikájának” modern kori elterjedésével a szoros olvasás módszere angolszász területen már a 19. század utolsó évtizedeitől növekvő népszerűségnek ör-

<sup>1</sup> Nathan K. Hensley, „Figures of Reading,” *Criticism* 54, 2. sz. (2012): 329–342, 340, <https://doi.org/10.1353/crt.2012.0013>.

vendett az általános és középiskolai oktatásban, valamint az egyetemeken.<sup>2</sup> Ez a módszer valójában egészen a 17–18. századi Biblia-értelmezésekig visszakövethető, sőt annál is messzebb, az antik retorikai gondolkodásig.<sup>3</sup> E retorikai és hermeneutikai hagyományokat követve a szövegek szoros vagy lassú vizsgálata a költészetéről szóló romantikus diskurzusban vett újabb lendületet, ahogy azt az „elsietett döntésekkel” (*rashness of decision*) szembeni wordsworthi intellemtanúsítja a *Lírai balladák* elé fűzött figyelmeztetésben. A Google Books Ngram Viewer által kínált grafikon az angol *close reading* kifejezés 1800 és 2000 közötti előfordulási gyakoriságára vonatkozólag – amint azt Michael Hancher bemutatta – alátámasztani látszik a fokozatos elterjedés tézisé,<sup>4</sup> közelebből nézve azonban mintha enyhén meredekebb emelkedést mutatna mind a második világháborút követően, mind az 1970-es évek végétől, ami az újkritika, illetve a dekonstrukció népszerűsödését jelezheti.



1. ábra. A *close reading* kifejezés előfordulási görbéje 1800-tól 2000-ig a Google Books Ngram Viewer<sup>5</sup> szerint (korpusz: English 2012)

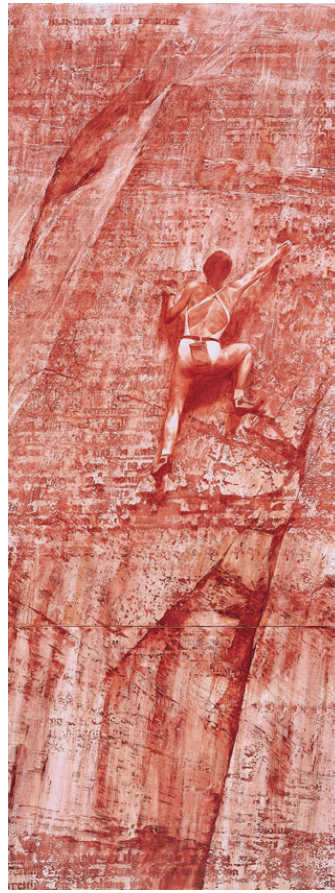
Végeredményben tehát a kifejezés közkeletű összekapcsolása az utóbbi tendenciákkal nem feltétlenül téves. Mark Tansey, kaliforniai művész, például aligha szerepeltette volna Paul de Man *Blindness and Insight* című 1971-es könyvét a *Szoros olvasás* (*Close Reading*, 1990) című festményén, ha az előző egy-két évtizedben a dekonstrukció nem járult volna hozzá elemi mértékben a szoros szövegelemzés gyakorlatának továbbfejlesztéséhez.

<sup>2</sup> Jay Jin, „Problems of Scale in »Close« and »Distant« Reading,” *Philological Quarterly* 96, 1. sz. (2017): 105–129, 105, 109.

<sup>3</sup> Michael Hancher, „Re: Search and Close Reading,” in *Debates in the Digital Humanities 2016*, eds. Matthew K. Gold and Lauren F. Klein (Minneapolis: University of Minnesota Press, 2016), 122–124, <http://dx.doi.org/10.5749/j.ctt1cn6thb.15>.

<sup>4</sup> Hancher, „Re: Search and Close Reading,” 127.

<sup>5</sup> Google Books Ngram Viewer, hozzáférés: 2020.07.02, <https://books.google.com/ngrams>.



2. ábra. Mark Tansey, *Szoros olvasás* (*Close Reading*, 1990)

Tansey vörösbarna monokróm képén egy szabadmászó nőalakot látunk, ahogy egy meredek sziklafalon kapaszkodik felfelé, melynek textúráját szó szerint textuális töredékek alkotják, konkrétan de Man tanulmánykötete 147. oldalának darabkái.<sup>6</sup> Ugyanez az oldal szerepel Tansey ugyancsak 1990-ben festett *Olvasó* (*Reader*) című képén, mely a „textuális abszorpció” egy másik jelentét, egy másik atletikus figura részben elmosódott alakját ábrázolja,<sup>7</sup> szintén hátulról, amint épp intenzív mozgásban van a szöveg felé, a szöveg belsejébe, vagy a szöveggel együtt. A két kép előtanulmányként is felfogható Tansey ugyanezen évben készült nagyszabású festményéhez, *A Grand Canyon építése* (*Constructing the Grand Canyon*) című alkotáshoz, mely nem csupán olyan posztstrukturalista vagy dekonstruktív olvasókat jelenít meg, mint de Man vagy Derrida, hanem a textuális terepnummal kapcsolatos viszonyulások és tevékenységek egész tárházát kínálja (mászás, lökés, fúrás, felmérés, kiaknázás, gyűjtés, üldögélés, járkálás stb.). A hegymászó „szoros olvasó” nőalakja biztonsági heveder nélkül jelenik meg, aki vakmerőn, kizárólag szövegszerű bizonyítékokra próbálja alapozni olvasatát,

<sup>6</sup> Marc Redfield, „Courses Taught by Paul de Man during the Yale Era,” in *Legacies of Paul de Man*, ed. Marc Redfield (New York: Fordham University Press, 2007), 177, <http://doi.org/10.5422/fordham/9780823227600.001.0001>.

<sup>7</sup> Mark C. Taylor, *The Picture in Question: Mark Tansey and the Ends of Representation* (Chicago: The University of Chicago Press, 1999), 80, <http://doi.org/10.7208/chicago/9780226791272.01.0001>.

kiszögellésekbe és repedésekbe kapaszkodva, vagy ezeken igyekezve megvetni a lábát, miközben a magasból, a kanyon pereméről két bölény méla könnyel bámulja a sokféle szöveg munkát megjelenítő színteret.<sup>8</sup>

Ahogy a *Blindness and Insight* tanúsítja, de Man mélyen le volt kötelezve a szoros szövegolvasás amerikai gyakorlatának. Méltányolta „az irodalmi kifejezőmód leheletfinom részleteinek leírásában mutatott figyelemreméltó kifinomultságot”, mint ahogy a „forma” egységére helyezett hangsúlyt, és az ezzel járó „érzék a kontextus iránt, ami a francia vagy német értelmezésekből gyakran hiányzik”, miközben komoly fenntartásai voltak az újkritikusok által hirdetett organicista és ahistorikus formalizmussal szemben.<sup>9</sup> Ugyanakkor éppígy kétségei voltak a kortárs német és francia irodalomkritika eredményei iránt is, mivel azokból hiányolta a szövegek apró részletei iránti figyelmességet. Éppen ezért üdvözölte annyira Jacques Derrida példaszerűen szövegközeli Rousseau-olvasatát,<sup>10</sup> s ezért szolgáltatót később maga is számos példát az efféle szigorú szövegelemzésre. Szélsőségig hajtva, az ilyen olvasás „a betű prózai materialitásához” vezet,<sup>11</sup> ami azonban nem sorolható be az empirista vagy fizikalista anyagfelfogás alá. Ahogy Derrida kommentálja: „A betű literalitása valójában nem azért szituálja ezt a materialitást, mert fizikai vagy érzéki (esztétikai) szubsztancia volna, hanem mert annak a prózai ellenállásnak a helye [...], amely mindenféle organikus vagy esztétikai totalizálásnak, minden esztétikai formának ellenáll”.<sup>12</sup>

Ma a szoros vagy lassú olvasás még mindig széles körben elterjedt gyakorlat az irodalomtudományon belül a retorikai olvasástól a pszichoanalitikus vagy gender szempontú elemzéseken át a posztkoloniális vagy különféle történeti érdekeltségű vizsgálatokig. A szoros olvasás ugyanakkor egyre inkább támadás alatt is áll, ezúttal nem a maradi erkölcsiség oldaláról, ahogy a 1980-as évek elején, hanem önnön formális ellenpárja, a „távoli olvasás” (*distant reading*) felől, a számítógépes segédlettel végzett elemzések irányából. (Az egyéb kapcsolódó módszertanok közül az „algoritmikus olvasás” vagy a „makroanalízis” volna még említhető.) Ahogy *Distant Reading* című 2013-as (ám mintegy két évtized tanulmányait újraközlő) könyvében a módszer legfőbb kifejlesztője és népszerűsítője, Franco Moretti fogalmaz, a távoli olvasás nem más, mint „azonosítani egy elkülönülő formai jegyet, majd nyomon követni, milyen átalakulásokon megy keresztül szövegek egész során át”.<sup>13</sup> Moretti örömmel fogadja Jonathan Arac meghatározását, miszerint a távoli olvasás voltaképp „szoros olvasás nélküli formalizmus”,<sup>14</sup> míg ő maga olykor inkább „szeriális” olvasásnak vagy „kvan-

<sup>8</sup> Tansey egyéb dekonstrukcióval kapcsolatos festményeiként lásd még a *Híd a karteziánus hasadék fölött* (*Bridge over the Cartesian Gap*), valamint a *Derrida vezeti de Man* (*Derrida Queries de Man*) című képeket (mindkettő 1990-ből való), melyek a *Blindness and Insight* 146., illetve 147. oldalát használják fel.

<sup>9</sup> Paul de Man, *Blindness and Insight: Essays in the Rhetoric of Contemporary Criticism* (New York: Oxford University Press, 1971), 27.

<sup>10</sup> De Man, *Blindness and Insight*, 110.

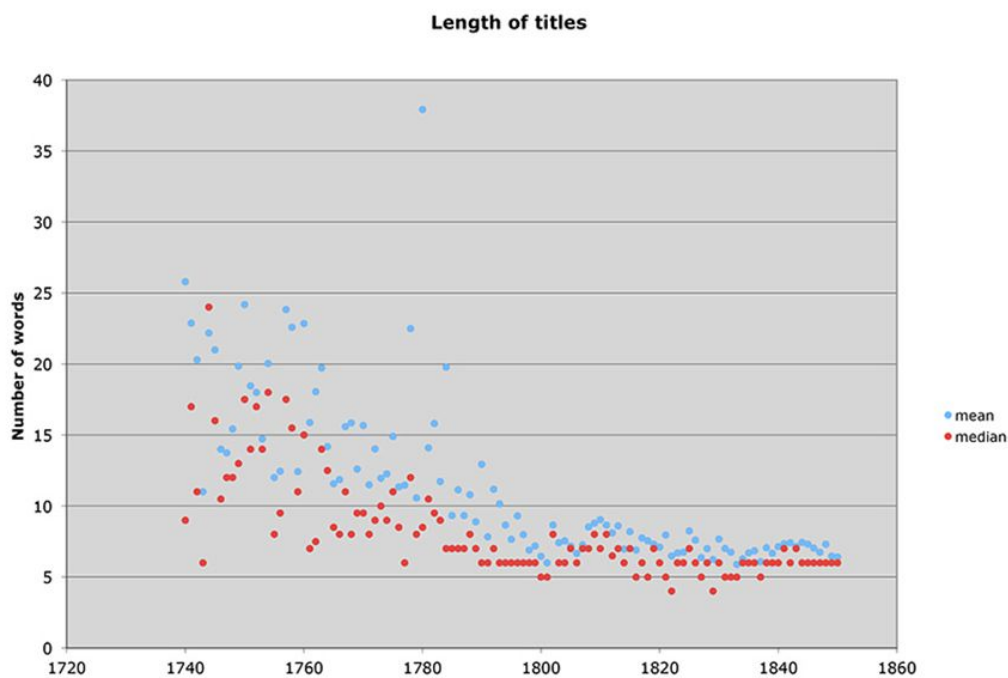
<sup>11</sup> De Man, *Esztétikai ideológia*, ford. Katona Gábor (Budapest: Janus/Osiris Kiadó, 2000), 79.

<sup>12</sup> Jacques Derrida, „Typewriter Ribbon: Limited Ink (2) (»within such limits«),” in Barbara Cohen, J. Hillis Miller, Andrzej Warminski and Tom Cohen, eds., *Material Events: Paul de Man and the Afterlife of Theory* (Minneapolis: University of Minnesota Press, 2001), 350.

<sup>13</sup> Franco Moretti, *Distant Reading* (London: Verso, 2013), 65.

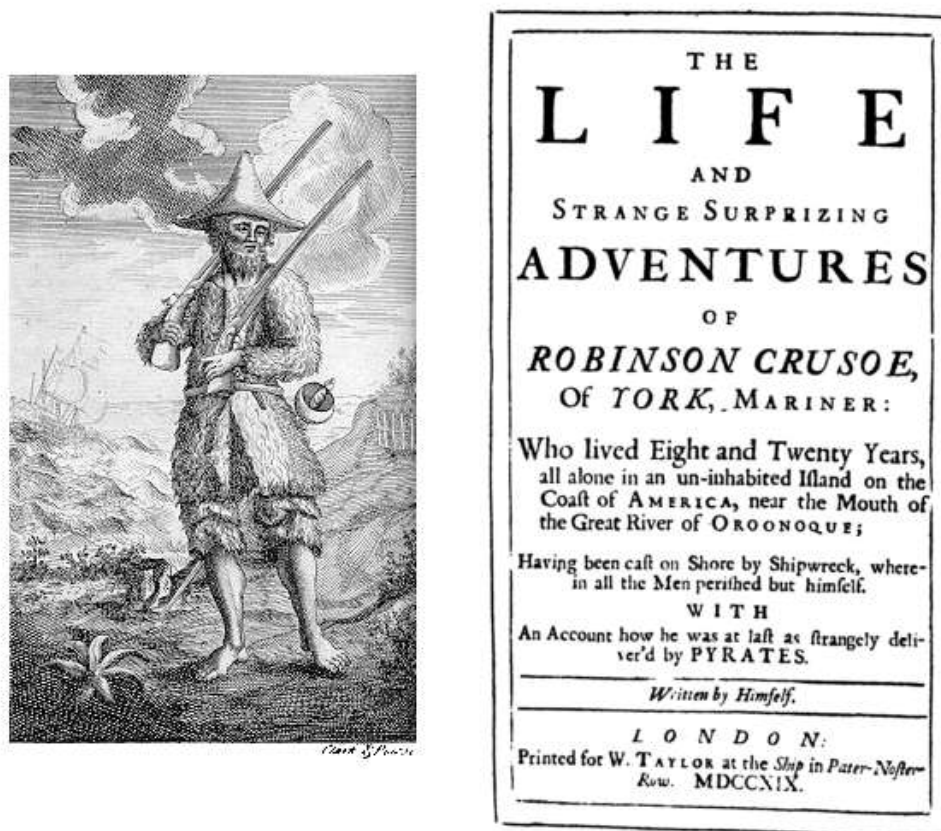
<sup>14</sup> Moretti, *Distant Reading*, 65.

titatív” analízisnek nevezi. Moretti szerint az „a baj a szoros olvasással”,<sup>15</sup> hogy egy (szűk) kánonra korlátozódik, s ezzel egy szekularizált teológiát éltet. A szövegközei vizsgálat azért kérdőjeleződik meg, mert érzéketlen olyan nagyobb mintázatokra, melyeket az emberi pillantás nem képes befogni. Az archívumok (a gondos kódolással előállított adatbázisok) számítógépes segédlettel megvalósuló elemzése ezzel szemben a mindeddig észrevétlen jelenségeket is láthatóvá teszi a „szerializálás” (*seriation*) által, amikor olyan hatalmi és ideológiai mintázatokat vizualizál, melyeket korábban soha senkinek nem volt lehetősége szemügyre venni. Egy 7000 egységből álló korpusz elemzésén keresztül a távoli olvasás képes kimutatni például, mennyire drasztikusan lerövidült a brit regények címe a 18. század során (például Defoe Robinsonjának hosszadalmas címéhez képest), ami Moretti szerint a piac intenzív növekedését bizonyítja.



3. ábra. A brit regénycímek rövidülése 1740-től 1850-ig (Moretti, *Distant Reading*, 183)

<sup>15</sup> Moretti, *Distant Reading*, 48.



4. ábra. Daniel Defoe regényének címe az eredeti címlapon: *Robinson Crusoe, yorki tengerész, élete és különös meglepő kalandjai; aki hosszan élt...* (1719)

A rövidség mellett két másik módját is bemutatja annak, hogyan is működnek a regénycímek tökéletes hirdetésként: stílusuk egyre metaforikusabb, ami felkelti a fogyasztók érdeklődését, továbbá cseles grammatikai formulákat alkalmaznak, melyek maguk is várakozásokat keltenek a narratív tartalmat illetően (például határozott vagy határozatlan névelők használatával, vagy birtokos szerkezetek alkalmazásával). Mindez szem előtt vész, ha az olvasók csupán a művek parányi csoportjának (a kánonnak) a tüzetes vizsgálatára szorítkoznak.

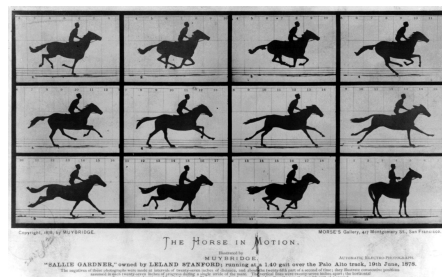
Az apró formai jegyekre helyezett folyamatos hangsúly azonban azt is jelenti, hogy a távoli olvasás korántsem próbál minden tekintetben túllépni a szoros szövegelemzésen. Inkább arra törekszik, hogy kiegészítse, méghozzá azzal, hogy az egyes textuális elemek aprólékos vizsgálatát kombinálja egy technikailag kiterjesztett érzékenységgel, mely immár nagyobb mintázatokot is észlel: „meg kell tanulnunk jelentést találni az *apró* változásokban és a *lassú* folyamatokban”.<sup>16</sup> Az „aprónak” a „lassúval” való kombinálása érdekében a „szoros” olvasásmódot a „gyors” (vagy „távoli”) olvasással kell összekapcsolni. Úgy is mondhatnánk: a bezúmolást kizúmolással, a „fentről-

<sup>16</sup> Moretti, *Distant Reading*, 192 – saját kiemelés.

le” vizsgálatot „lentről-fel” vizsgálatokkal kell ötvözni.<sup>17</sup> Pontosan ez az, amit Moretti képvisel, amikor olyan „kvantitatív vizsgálatot” sürget, amelynek „egységei nyelvi és retorikai elemek”.<sup>18</sup> A műfajok történelmi evolúciója esetében például az efféle kombinált elemzés a retorikai olvasás térbeli ráközelítéseit a kvantitatív elemzés időbeli madártávlatával egészítené ki. Megpróbálná technikailag időben lekicsinyíteni az érzékileg térben felnagyítottat.

## 2. Stanford

Hasonló ambivalencia jelenik meg Stanford (mint egyetem és mint alapító) kapcsán is, amennyiben a távoli és a közeli elemzésmódok keverékeként tekintjük. Míg a Stanford Egyetem sokunk számára a távoli perspektívát vagy a nagysebességű olvasást képviseli, mivel a Literary Lab kutatóközpontnak az otthona (mely máig a mostanra emeritusszá lett alapító, Moretti nevével kapcsolódik össze),<sup>19</sup> addig az egyetem 19. századi alapítója, az üzletember és politikus Leland Stanford épp ellentétes irányba tett lépéseket, és a közeli vizsgálatot segítő technológiák, a mozgást lelassító vagy akár ki is merevítő technikák kifejlesztéséből vette ki részét. Mint ismeretes, az ő megrendelésére fotózott ügető vagy vágta lovakat Eadweard Muybridge, akinek lélegzetelállító fotósorozatai (köztük az az 1878. június 19-én készült képsor, mely Stanford saját „Sallie Gardner” nevű versenylovát ábrázolja a Palo Alto-i versenypályán 60km/órás vágta közben)<sup>20</sup> fontos szerepet játszottak annak megváltoztatásában, ahogyan az állatok mozgását érzékeljük.



5. ábra. Eadweard Muybridge, „A ló mozgás közben” (Leland Stanford „Sallie Gardner” nevű versenylova vágta közben 1878. június 19-én a Palo Alto-i versenypályán)

<sup>17</sup> Stefan Jänicke, Greta Franzini, Muhammad Faisal Cheema and Gerik Scheuermann, „On Close and Distant Reading in the Humanities: A Survey and Future Challenges,” in *Eurographics Conference on Visualization*, eds. Rita Borgo, Fabio Ganovelli and Ivan Viola (The Eurographics Association, 2015), 10–11, <http://doi.org/10.2312/eurovisstar.20151113>.

<sup>18</sup> Moretti, *Distant Reading*, 204.

<sup>19</sup> Lásd Moretti cikkeit (több ízben társszerzőként) a 2010-es alapítású Stanford Literary Lab 2011-ben indult *Pamphlets* sorozatában, hozzáférés: 2020.02.10, <https://litlab.stanford.edu/pamphlets/>.

<sup>20</sup> Eadweard Muybridge, „The Horse in Motion (»Sallie Gardner«, owned by Leland Stanford; running at a 1:40 gait over the Palo Alto track, 19th June 1878),” Library of Congress Prints and Photographs Division, Washington, D.C., hozzáférés: 2020.02.10, <http://loc.gov/pictures/resource/cph.3a45870/>.



Bár az említett képsor éppenséggel vágát ábrázol, Stanford eredetileg azért rendelte meg ezeket a fényképsorozatokat, hogy fotografikus bizonyítékot kapjon arra a meggyőződésére, miszerint ügetés közben is vannak olyan pillanatok, amikor a ló négy lába közül egyidejűleg egyik sem érinti a talajt. Muybridge technikai megoldása a korai akciófotózás tartós eredményének bizonyult. A képek elkészítésére összerakott kifinomult felvevőrendszer két sor fényképezőgépéből állt, melyekből 24 darab 90 fokos szögben a pálya mentén, további 12 darab pedig 30–60 fokos szögben a pálya végén lett elhelyezve 1/2000 másodperces záridővel úgy, hogy a készülékeket kioldózsínók (később óraszerkezet) hozták működésbe, ahogy a ló haladt. Az ekként készített képek lehetővé tették az állati helyváltoztató mozgás különböző fázisainak tanulmányozását a maguk elszigetelt kimerevítettségében, és kellő bizonyítékot szolgáltatottak Stanford kezdeti feltételezésének alátámasztására. Muybridge azonban animálta is a sorozatait, méghozzá egy maga alkotta készülékkel, az úgynevezett „zoopraxiszkóppal”,<sup>21</sup> mely nemcsak hogy újra mozgásba hozta a pillanatképeket (előremozdítva ezzel a mozi-technika fejlődését), hanem lehetővé tette a lovak mozgásának lelassított vizsgálatát is. Muybridge vágázó lováról készült kimerevített vagy animált fotósorozatai a vágtaábrázolás hosszú festészeti hagyományának kimozdításához is hozzájárultak. Ahogy George Stubbs alkotásain (például a *Baronet* című 1791-es festményén), vagy Erdmann Gottlieb Prestelnek *Az első pesti lóversenyt* (1827) ábrázoló képén megfigyelhető, a festészet hagyományos darabjai a vágázó lovakat még előrenyújtott első és hátranyújtott hátsó lábakkal mutatták.



6. ábra. George Stubbs, *Baronet* (1791)

<sup>21</sup> Muybridge zoopraxiszkópja nem volt teljesen új minden elemében: hasznosított olyan korábbi készülékeket, mint a fenakisztiszkóp (1832) és a zoetróp (1833). A munka technikai részleteiről lásd Muybridge könyvét: *Descriptive Zoopraxography* (1893). Muybridge később más házi és vadállatokról is készített fotósorozatokat, sőt mindennapi tevékenységet végző vagy sportoló emberekről is. Képeinek számos gyűjteménye közül lásd az alábbi modern kiadást: *Muybridge's Complete Human and Animal Locomotion: All 781 Plates from the 1887 Animal Locomotion* (New York: Dover, 1979).



7. ábra. Erdmann Gottlieb Prestel, *Az első pesti lóverseny* (1827)

Muybridge fotósorozata örökre megváltoztatta ezt a képet, felfedte ugyanis, hogy a ló lábai még vágta közben is minden egyes pillanatban más pozícióban vannak egymáshoz képest, így a kinyújtott lábak helyére idővel egy másik kép lépett, mely azóta a westernfilm-plakátok emblematikus elemévé vált, abban a pillanatban mutatva a vágató lovat, amikor a teste alatt a levegőben összehúzza lábait. A Stanford által megrendelt fényképek így aztán döntő szerepet játszottak annak előhozatalában, amit Walter Benjamin később „optikai tudattalannak” nevezett. Az ügetés vagy vágózás részleteinek láthatóvá tételével egy olyan jelenség vált megfigyelhetővé, amely a szemünk előtt volt ugyan, mégis rejtve maradt mindaddig, amíg a (kimerevített vagy lelassított) fotografikus-mozgóképi ábrázolás hozzáférhetővé nem tette a tüzetesebb temporális vizsgálat számára. Mindez némi habozásra készíthet bennünket a tekintetben, hogy melyik hagyományba is soroljuk Stanfordin (az alapítót és az alapítottat): hogy vajon a gyors, szeriális elemzés élharcosának tekintsük-e, vagy a lassítás kifejtőjének. Annyi mindenesetre bizonyos, hogy ezek a technikai tendenciák eleve nem is „két” külön hagyományt alkotnak.

### 3. Benjamin

Az „optikai tudattalannal” Walter Benjamin arra utalt, ahogy a mozgóképtechnika segítheti olyan térbeli és időbeli események megfigyelését, amelyeket az emberi szem képtelen lekövetni. „A műalkotás a technikai reprodukálhatóság korában” (1939) című tanulmányának egyik gyakran idézett bekezdésében azt írja:

S ekkor jött a film, és a tizedmásodpercek dinamitjával szétrobbantotta ezt a börtönvilágot. Így a szétszórt romok közt most már bizvást vállalkozhatunk kalandos utazásokra. A közelkép [*Großaufnahme*] kitágítja a teret, a lassítás [*Zeitlupe*] pedig az időt. S ahogy a nagyításnál [*Vergrößerung*] korántsem az

„amúgy is” pontatlanul látható dolgok pusztá magyarázásáról van szó, hanem sokkal inkább az anyag teljesen új szerkezeti képződményei kerülnek előtérbe, úgy a lassítás [*Zeitlupe*] sem csupán ismert mozgásmotívumokat láttat, hanem a már ismertben valami teljesen ismeretlent tár fel. [...] Itt avatkozik be a kamera a maga segédeszközeivel: a buktatással és az emeléssel, a megszakítással és izolálással, a folyamat megnyújtásával [*Dehnen*] és sűrítésével [*Raffen*], a nagyítással [*Vergrößern*] és kicsinyítéssel [*Verkleinern*]. Az optikai tudattalanról csak rajta keresztül szerzünk tudomást, ahogyan az ösztönös tudattalanról a pszichoanalízis segítségével.<sup>22</sup>

Bár nyilvánvaló történelmi és technikai okokból kifolyólag Benjamin elsősorban a térbeli és időbeli nagyításra – a közelképre (*Großaufnahme*, „nagy felvétel”) és a lassításra (*Zeitlupe*, „időnagyító”) – koncentrált, a fenti passzus végén jelzi a másik irányt, a képek gyorsításának vagy kicsinyítésének lehetőségét is a folyamat „sűrítésével” vagy a tárgy „kicsinyítésével”. Moretti madártávlati beállításai Benjamin tervezetének e második felét látszanak beteljesíteni. Ha Benjamin ma élne (egy olyan korban, mely egyszerre a superlassítások és a „time-lapse” videók kora), alighanem Párizsból Kaliforniába helyezné át a *Passzázsok* (*Passagen-Werk*) fáradtságos adatbázis-fejlesztési munkálatait – amely mű egyébként maga is egyfajta társadalmi-gazdasági irányultságú „távoli olvasás” *avant la lettre* –, és most ő volna a Literary Lab vezetője a Stanfordon, talán épp Franco Morettivel mint tudományos segédmunkatárssal az oldalán. (Moretti amúgy is elismerően utalgat Benjaminra, igaz, sehol nem szentel neki érdemi figyelmet. Lásd például, hogyan használja fel Benjamin értelmezését Baudelaire versengő poétikájáról, vagy lásd a szórakozás benjaminini témájára tett utalását.<sup>23</sup>)

A technika fejlődésével és a digitális bölcsészet megállíthatatlan előretörésével az irodalomtudomány fogalma, vagy magáé az olvasásé, ma talán éppúgy traumatikus átalakuláson megy keresztül, mint jó száz évvel ezelőtt a fényképezés és a film megjelenésével a festészet és színház hagyományos művészeti ágainak mintájára elgondolt művészeté – ahogy arról a Valérytól vett mottóban olvashatunk a Benjamin-esszé harmadik változatának elején. E változás traumájának egyik leglátványosabb tünete maga az, ahogyan egyesek tagadják, hogy a „távoli olvasás” egyáltalán olvasási aktus volna. Így tesz például Jonathan Culler, aki szerint Moretti gyakorlata „egyáltalán nem is olvasás”.<sup>24</sup> Culler a kultúrpeszimizmus tónusában ad hangot balsejtelmeinek egy olyan kort illetően, ahol „az új elektronikus források lehetővé teszik, hogy tényleges olvasás nélkül végezzünk irodalmi kutatást”.<sup>25</sup> Ezzel függ össze Culler azon meggyőződése, hogy a szoros olvasás valódi ellenpólusa – minden látszat ellenére – nem a

<sup>22</sup> Walter Benjamin, „A műalkotás a technikai reprodukálhatóság korában,” ford. Kurucz Andrea, átdolg. Mélyi József, hozzáférés: 2020.02.10, [http://aura.c3.hu/walter\\_benjamin.html](http://aura.c3.hu/walter_benjamin.html); „Das Kunstwerk im Zeitalter seiner technischen Reproduzierbarkeit (Dritte Fassung),” in *Gesammelte Schriften*, Bd. 1., Hrsg. Rolf Tiedemann und Hermann Schweppenhäuser (Frankfurt am Main: Suhrkamp, 1991), 1: 500.

<sup>23</sup> Moretti, *Distant Reading*, 70, 174.

<sup>24</sup> Jonathan Culler, „The Closeness of Close Reading,” *ADE Bulletin* 149. sz. (2010): 20–25, 20, <http://doi.org/10.1632/ade.149.20>.

<sup>25</sup> Culler, „The Closeness of Close Reading,” 24.

távoli olvasás (mely, mint épp az imént láttuk, egyáltalában nem is számít olvasásnak), hanem a „hanyag” vagy „laza” olvasás, az olyan, amelyik elsietett leegyszerűsítéseket tesz.<sup>26</sup> Bár ezt az elgondolást mások is magukévá tették,<sup>27</sup> és további érvekkel támasztották alá (például a sűrűsége mint a közelség „valódi” értelmére utalva),<sup>28</sup> a törekvés, mely azzal próbálja semlegesíteni a távoli olvasást, hogy nem is tekinti valódi riválisnak, számomra kissé elhamarkodottnak tűnik, mivel elfelejt kritikusan számot vetni azzal a hosszú metafizikai hagyománnyal, amely a gyorsaságot (vagy a távoli pillantást) a jelentéktelenséggel és felszínességgel társította. Egy másik visszatérő kijelentés, melyet érdemes megemlíteni, a „szcientizmus” miatti aggodalom. Ez a félelem messzemenően jogosnak tűnik, amennyiben olyan kritikai gesztusként értékeljük, mely a természettudományos diskurzus fetisizált státusát veszi célba, ugyanakkor megfélelkezni látszik e diskurzus és az azt övező aggályok történetiségéről, leginkább is arról, hogy néhány évtizeddel ezelőtt magát a szoros olvasást érezték sokan túlságosan tudományosnak vagy technikainak, s tekintették ezért fenyegető veszélynek arra nézve, ami *akkoriban* az olvasás egy természetesebb módjának tűnt.<sup>29</sup> A tagadás efféle gesztusai, amiként az „olvasás körüli háborúskodás” kirobbanása is, jól bizonyítják, micsoda léptékű változások zajlanak ma az olvasás technikái és technológiai, s ezek hátterében a „természetesség” különféle felfogásai körül. E változások nyomán, úgy tűnik, maga az olvasás fogalma is alapvető átalakuláson megy keresztül.

Ezzel együtt a kép nem túlságosan kedvező a szorosan olvasó kutatókra nézve. Egy olyan korban, amely sokszor „fétissé” alakítja a nagy időfeszítávokat,<sup>30</sup> nem sok babér jut az aprólékos szövegelemzésnek. A távoli perspektívák hozzáadása nélkül a szövegközeli elemzés vak marad a nagyléptékű történelmi folyamatokra. Az olyan szoros olvasók, mint de Man, soha nem lesznek képesek meglátni a fától az erdőt, és miként az újdonsült játékos a térképes játékban, *Az ellopott levél* című Poe-novellában, arra vannak ítélve, hogy lúzerek és nevetség tárgyai legyenek az agyafúrt szeriális elemző (Moretti) szemében, akinek ellenben távlatos perspektívája van, így képes megpillantani az irodalom társadalmi-gazdasági dimenzióit és a történelem évtizedeken vagy akár évszázadokon átívelő rejtett folyamatait.

#### 4. Poe

Poe elbeszélésének említett részlete megfelelő kiindulópontul szolgálhat, ahonnan talán még ennél is tovább léphetünk. Poe különc nyomozója, Auguste Dupin, így írja le a térképes játékot:

– Van egy kitalálós játék – folytatta Dupin –, térképen játsszák. Az egyik fél felad a másiknak egy szót, egy folyó, város, állam vagy birodalom nevét, hogy keresse meg a térkép tarkabarka, zűrzavaros felületén. Aki kezdő ebben a játékban, az rendesen azzal akarja zavarba ejteni ellenfelét, hogy a legapróbb betűkkel

<sup>26</sup> Culler, „The Closeness of Close Reading,” 20.

<sup>27</sup> Hancher, „Re: Search and Close Reading,” 125; Jin „Problems,” 106, 118–120.

<sup>28</sup> Hancher, „Re: Search and Close Reading,” 124.

<sup>29</sup> Jin, „Problems,” 118–120.

<sup>30</sup> Eric Hayot, „A Hundred Flowers,” in *Graphs, Maps, Trees: Responses to Franco Moretti*, eds. Jonathan Goodwin and John Holbo (Anderson: Parlor Press, 2011), 66.

nyomott nevet keresi ki, a haladó azonban olyan szavakat választ, amelyek nagy betűkkel, a térkép egyik szélétől a másikig érnek. Mint a túlságosan nagy betűs cégtáblák és hirdetések az utcán, ezek éppen nagyságuk [*being excessively obvious*] miatt kerülnek el a figyelmet, és itt a fizikai „meg nem látás” [*oversight*] teljesen rokon a pszichikai figyelmetlenséggel [*inapprehension*], amelynek következtében elhaladunk túlságosan is látható és magától értetődő dolgok mellett, és azoknak semmiféle figyelmet nem szentelünk [*the intellect suffers to pass unnoticed those considerations which are too obtrusively and too palpably self-evident*].<sup>31</sup>

Az elbeszélés logikája szerint a rendőrfőnök mikroszkopikus kutakodása azért van eredménytelenségre ítélve, mert az ellopott levelet épp saját szembeszökő, nyilvánvaló vagy magától értetődő volta rejti el. A levél egyáltalán nincs eltakarva. Épp ellenkezőleg, ki van emelve, sőt túlságosan is előtérbe van állítva, „túl szembeszökő” (*too palpable*) és „túl nyilvánvaló” (*too obtrusive*). Poe elbeszélése e ponton a figyelem mint fiziológiai vagy mentális fókuszálás burkolt elméletét kínálja, hiszen Dupin a térképes játék leírását azzal a kérdéssel vezeti fel, hogy „melyik cégtábla kelti fel legjobban a járókelők figyelmét [*which of the street signs over the shop-doors are the most attractive of attention*]” (56; 222).

A figyelem tényleg meghatározó szerepet játszik a történet korábbi és későbbi részében. Döntő szerepe van a levél ellopásának ősjelenetében, mely az illető cselekményt nem is egyszerűen lopásként mutatja, hiszen félig nyílt tettről van szó, melyet a levél jogos tulajdonosának szeme láttára követ el a „hiúzszerű” miniszter, és amely csakis a harmadik személy elől marad rejtve, aki számára tartalma botrányos volna (44; 207). Ha a lopás rejtett, de békés tett, míg a rablás nyílt, de erőszakos, akkor az „ellopott” levél – a jogos tulajdonos vonatkozásában – bizony inkább egy mesteri rablás, semmint trükkös lopás által vétetik el, hiszen a tulajdonost nem átverték, hanem rákényszerítették, hogy engedje át a levelet, ha nem akarja, hogy a harmadik személy észrevegye. Mindent egybevéve, az észrevevések, tekintetek és észre nem vételek eme hálójában a levél félig ellopott, félig elrabolt portéka. A figyelem hasonlóképp döntő szerepet játszik a levél visszalopásában is (mely immár csakugyan tisztán lopás). Ezt az aktust Dupin a miniszterhez tett első látogatásakor alapozza meg (az eredeti levél aprólékos, de rejtett megfigyelésével), a második látogatás során pedig a miniszter figyelmének pillanatnyi elvonásán keresztül teszi lehetővé (egy jól megkoreografált utcai incidenst hívva segítségül).

Minthogy Poe elbeszélése a figyelem jeleneiből (a figyelem odavonásának vagy elvonásának, keresésének vagy elkerülésének pillanataiból) van összeszöve, amit kapunk, az nem egyszerűen a figyelem általános elmélete, hanem speciálisabban a figyelemkeresés és a figyelemkerülés párhuzamos elmélete.<sup>32</sup> A túlméretezett utcai reklámok példája („a túlságosan nagy betűs cégtáblák és hirdetések az utcán”) azt mutatja

<sup>31</sup> Edgar Allan Poe, „Az ellopott levél,” ford. Pásztor Árpád, in Edgar Allan Poe, *Túl életem és halálom* (Budapest: Népszava, 1989), 56; „The Purloined Letter,” in Edgar Allan Poe, *Tales of Mystery and Imagination* (Ware: Wordsworth Editions, 1993), 222 (további hivatkozások – ugyanebben a sorrendben – a főszövegben).

<sup>32</sup> Arról, hogy Poe elbeszélése miként illeszthető be a figyelem egy általánosabb elméletébe, másutt írtam (Dupin Campanellára tett utalása kapcsán): „Teletrauma: Distance in Burke’s *Philosophical*

be, hogyan jöhetnek kapóra a figyelemkeresés sikertelen kísérletei akkor, amikor épp a „túlzott nyilvánvalóság” (*being excessively obvious*) általi figyelemelkerülés a cél. A „fizikai meg nem látás” (*physical oversight*) és a „pszichikai észre nem vevés” (*moral inapprehension*) közti analógiával az észrevétel elkerülésének ajánlott taktikája a tiszta vizualitás világából a mentális műveletek széles terepére helyeződik át, és tökéletes receptül szolgál bárkinek, aki el akar rejteni valamit a rendőrségi nyomozás elől. A tolvaj és a rendőrség közti térképjátékban, úgy tűnik, minden a kutatás „körén” (*range*), vagyis a léptéken múlik, ahogy azt egy korábbi passzusban Dupin szó szerint meg is jegyzi: „Ha a levél a kutatási körükbe [*range*] esik, a fickók kétségtelenül meg is találják” (50; 214). A levél nem lett eldugva semmi más mögé, de nem is lett összetömörítve szinte a megsemmisítésig. Pont ellenkezőleg: az előtérben hagyták, sőt túlságosan is előtérbe állították. Ezért Dupinnek ki kell zúmolnia a képből, hogy észrevegyen valami olyat, ami minden elrejtés nélkül is elkerülte volna a figyelmét. A tolvaj furmányossága ugyanis épp abban rejlik, hogy azzal próbálta „elrejtetni ezt a levelet” (*to conceal this letter*), hogy „meg sem próbálta elrejtetni” (*not attempting to conceal it at all*) (56; 222–223).<sup>33</sup>

A szöveg egyik passzusában, mely a kettős végtelen Pascal, Voltaire vagy Kant általi tárgyalásaira emlékeztet (vagy ha már itt tartunk, *Az olvasás allegóriáinak* Pascaltól vett mottójára: *Quand on lit trop vite ou trop doucement on n'entend rien*),<sup>34</sup> Poe zseniális detektívje kifejezetten fel is hívja a figyelmünket a vég nélküli tévelygés kockázatára, mely olyan jól látszódik a rendőrség ügyetlenkedésén. Ez a tévelygés lappangó fenyegetést jelent bárkire, aki „túlságosan alapos” (*too deep*) vagy „túlságosan felületes” (*too shallow*) módon állna neki a keresésnek (50; 215), mivel mindkét eljárás tévúton jár, ha eltéveszti a léptéket. Dupin példaként a legendás diákot említi, aki a „páros, páratlan” kitalálós játékban volt csodálni valóan sikeres. E példa kapcsán az ellenfél „ravaszságának” megfigyelését és felmérését hangsúlyozza, továbbá annak alapján a helyes lépték meghatározásának szükségességét. Elmagyarázza, miért nem vezet eredményre, ha a nyomozásban „nincs elvi változatosság” (*no variation of principle*) és mereven ragaszkodunk „egyetlen keresési elv vagy elvrendszer” (*one principle or set of principles of search*) érvényesítéséhez: „Nyomozásukban nincs elvi változatosság [...] Hogy valójában mennyit ér és mit jelent az a sok fúrás-faragás, méricskélés, nagyítóüveggel való vizsgálgatás, az épület felületeinek négyzethüvelykre való bontása – micsoda mindez, ha nem pusztán kibővítése *egyetlen keresési elv vagy elvrendszer alkalmazásának*, melyhez a rendőrfőnök hosszú pályafutása során hozzászokott?” (vö. 52; 216–217; kiemelés az eredetiben).<sup>35</sup> Konkrétan az ellopott levél esetében Dupint

*Enquiry*,” *The AnaChronisT* 17 (2012): 1–26 (lásd különösen a „Tension and Attention” című fejezetet). A figyelem általános problematikája mellett a figyelemkeresés és figyelemelkerülés kérdéséről is szól alábbi tanulmányom: „Figyelemcélpontok: egy idióma viszonya a terrorizmushoz és a technikához,” *Különbség* 1 (2017): 105–31.

<sup>33</sup> A fordítást módosítottam.

<sup>34</sup> „Amikor túlságosan gyorsan vagy túlságosan lassan olvas az ember, nem ért semmit”, lásd Blaise Pascal, *Gondolatok*, ford. Pődör László (Budapest: Gondolat Kiadó, 1983), 30 (69. töredék); franciául lásd: Blaise Pascal, *Pensées*, hozzáférés: 2020.02.10, <http://www.penseesdepascal.fr/XXV/XXV75—moderne.php> (75. töredék). Vö. Paul de Man, *Az olvasás allegóriái: Figurális nyelv Rousseau, Nietzsche, Rilke és Proust műveiben*, ford. Fogarasi György (Budapest: Magvető Kiadó, 2006), 5.

<sup>35</sup> A fordítást módosítottam.



megfigyelései és mérései arra a következtetésre vezetnek, hogy a tolvaj gondolkodásmódja nem csupán matematikusra, hanem költőre is vall, ami az elrejtés fentebb leírt kifinomultabb, sőt egyenesen paradox módját jelenti, mely a levelet épp az el nem rejtéssel próbálja elrejtetni. Ahhoz, hogy felérjen ellenfelével, Dupinnek analitikussá kell válnia, analitikussá viszont csakis akkor válhat, ha ő maga is „éppen annyira költő, mint matematikus”. Önmagában a matematikai okoskodás nem elég. Ki kell egészülnie a „fogások” (*ruses*), azaz fortélyok vagy cselek iránti költői érzékkel. Ahogy Dupin rámutat: „ha a miniszter csak matematikus lett volna, a rendőrfőnök nem szorul rá, hogy a csekket a nevére kiállítsa. De én tudtam, hogy D. éppen annyira költő, mint matematikus [*both mathematician and poet*] [...] Az, hogy olyan sokszor kimaradt éjszakára, amit a rendőrfőnök a siker zálogának tekintett, szerintem szintén csak egyik fogása [*ruses*] volt” (55; 220). Ezért dönt végül is Dupin a kizúmolás mellett, és ez az oka annak is, miért fordul elbeszélőként végül egy másik példához, a térképes játék példájához, amikor döntését magyarázni vagy illusztrálni szeretné. E ponton ritka, de annál emlékezetesebb módon reflektál arra, hogyan képes a retorika (pontosabban a képes beszéd: „a metafora vagy a hasonlat”) „megerősíteni” vagy „széppé tenni” a beszédet (55; 221).

Ám bármennyire fantáziadús és frappáns is a térképes játék példája Dupin szónoklatának végső díszeként, és bármennyire összhangban is van a léptékek váltogatására vonatkozó érvelésével, nem teljesen illeszkedik az általa előadott narratívához, pedig végső soron épp azt kellene egybefognia. A figyelem különböző léptékeinek fókuszba állítása ugyanis nem adja vissza híven a történet minden részletét. A megtalálás csúcspontját tekintve a térképes játék példája kissé félrevezető, az ellopott levél ugyanis nem egyszerűen csak triviális mérete és elhelyezkedése (viszonylagos nagysága és „túl kiugró helyzete” [*hyper-obtrusive situation*])<sup>36</sup> miatt kerüli el a rendőrség figyelmét. Megtálalásának roppant nehézsége a levél figurális elváltoztatásával is összefügg. Bár mérete változatlan, kinézete „egészen más” (*radically different*) (57; 224). Továbbra is levélnek néz ki, az igaz, de hogy miféle levélnek, az már egy másik kérdés. A nagy méret vagy a szembeszökő elhelyezkedés csupán egy olyan elrejtetlen tárgy illusztrációja, amely fölött minden láthatósága ellenére is elsiklik a tekintet, egyszerűen mivel a keresés léptéke nem illeszkedik a tárgy elhelyezésének léptékéhez. A levél azonban egészen más módon is kibújik a vizslató szemek szorításából: figurális átalakulással, azzal, hogy nem annak mutatkozik, ami. Miként Dupin hangsúlyozza, „a levelet kifordították, mint egy kesztyűt, aztán újra összehajtogatták, újra címezték, és lepecsételték” (58; 225). Sőt, félig még ketté is tépték, hogy azt a benyomást keltse, mintha csak egy közönséges fecni volna, egy minden különösebb jelentőség nélküli gyűrött dokumentum. Így az, hogy jól látható helyre teszik, inkább csak ráadás, túlbiztosítás, vagy egyszerű fricska ahhoz képest, hogy figurálisan átalakul. Ahhoz, hogy megértse a tolvaj költői vénáját, Dupinnek nem csak a lépték alapos beállítására kell ügyelnie, hanem rendelkeznie kell egy radikálisan „költői” érzékkel is a figuráció iránt, mely érzék ebben az esetben transzfiguráció, sőt diszfiguráció iránti érzék is kell legyen. Nem keresheti egyszerűen ugyanazt az „alakot” egy másik „mennyei” szinten. Ha ugyanis a költők ezt csinálják, akkor nem is igazán különböznek a mate-

<sup>36</sup> A magyar fordításban itt azt olvassuk: „olyan szembeötlően tovakodott elő rekeszéből” (57), vö.: „The Purloined Letter”, 224.

matikusoktól, legalábbis amennyiben a matematikát – Dupin meghatározása alapján – az „alak és mennyiség” (*form and quantity*) tudományának tekintjük (53; 219). Ha létezik bármi is, amit a költői szem láthat, a matematikus szeme viszont nem, akkor az nem a „forma” (bármilyen „kvantitásban”, méretben vagy számban forduljon is elő), hanem a deformáció vagy transzformáció, az eltorzulás vagy átalakulás. Ha a levelet megtévesztően átváltoztatták, akkor a felfedezéséhez nem csupán a különböző léptékek (vagyis ugyanannak a formának a különböző méretei) iránti költői érzék szükséges, hanem az elváltoztatás, a retorikai értelemben vett *permutatio* (tehát az allegória vagy ironia, avagy a figurális színlelés bármely más módja) iránti érzék is, ha egyáltalán van értelme ezekkel kapcsolatban még költői „érzékről” beszélnünk. Dupin költészetfelfogása még mindig túlságosan is matematikai, és az általa hirdetett másféle keresési elv még mindig nagyon is megegyezik azzal, amelyet túlhaladni szeretne, hiszen – hogy az egyik korábbi analógiára utaljak – az „alapos” és a „felületes” (avagy a „mély” és a „sekély”) egyazon rendszerhez tartoznak, s így ugyanannak a végzetnek vannak kitéve. Az, hogy a mikroszkopikus kutatás nem vezet eredményre, még nem jelenti azt, hogy a makroszintű keresés feltétlenül sikeresebb lesz.

Nem a kutatási léptéken *belül* kellene váltani (a közeliről a távolabbi vizsgálódásra, vagy fordítva), hanem a léptékről *magáról* kellene váltani valami másra, az érzékelésről az olvasásra, a szó szerinti jelentésről az átvitt értelemre, a formáról a deformációra, vagyis magában a keresési heurisztikában volna szükség radikális váltásra. Az olvasásnak nyomon kell követnie azt, amit meg vagy fel akar találni. A nyomozás pedig változást feltételez. A levél radikálisan léptéken kívüli, mivel semmilyen konkrét lépték nem garantálhatja megtalálását. Így aztán semmilyen távoli pillantás sem találhatja meg, nemcsak a közeli mustra. Dupin sikere csak részben köszönhető a távlatos perspektívának. Mert amit a fák helyett látnunk kellene, az még csak nem is az erdő, hanem egy olyan alak vagy mintázat a vadonban, mely csupán halványan hasonlít akár fákra, akár erdőre, s előfordulhat, hogy egyáltalán még a vegetációhoz sincs semmi köze. A 19. században oly népszerű képrejtvények, a bokrokban vagy a fák lombjai közt rejtőzködő (emberi, állati vagy élettelen) alakok, hasonlóan modernebb leszármazottaihoz (amilyen Voltaire kétértelmű büsztje Salvador Dalí emblematisztikus festményén, vagy amilyenek a rejtőzködő koponyák Orosz István grafikáin), mind-mind egy olyan „költői” foglalatosság példái, mely nem egyszerűen csak a léptékekkel operál, hanem a képlékenységgel, eltorzulással, anamorfózissal, vagy változékonysággal is. Amikor például, ahogy az iméntiekben Poe szövegével tettük, újraolvasunk egy művet, annak bizonyos passzusait, s tesszük ezt az illető mű bizonyos mértékű sorozatba állításával (Stanford, Benjamin, Poe...), akkor nehéz eldönteni, hogy ezúttal „közelebbről” vettük-e szemügyre, vagy éppen „távolabbról”, s épp a komparatív szerializálással hoztuk-e elő belőle egy olyan mintázatot (a közelítés/távolítás mintázatát), mely másként aligha lett volna látható, de legalábbis aligha nyert volna ilyenén értelmet vagy kritikai jelentőséget. Látható ugyanakkor az is, hogy a sorozatot alkotó elemek közti nem-azonos ismétlés miféle elcsúszásokat, hangsúlyáthelyezéseket, vagyis miféle mutációt foglal magában, s hogy a szövegegyüttesben felfedezhető mintázat kirajzolódásához mennyire szükség volt arra is, hogy e változásokra is tekintettel legyünk, vagyis hogy



a sorozat elemei közti ismétlődést másként-ismétlődésként, azaz iterációként fogjuk fel, mely minden komparatív elemzés alapja.<sup>37</sup>

A szoros és a távoli olvasás békés egymásmelletti és együttműködésén alapuló „rugalmas” vagy „zúmolható” olvasásmódok<sup>38</sup> színrelépésével az olvasás körüli „háborúskodás” egyszer talán a múlté lesz.<sup>39</sup> Egy nemrégiben megjelent tanulmányukban a Stanford számítógépes elemzői (köztük Moretti) rámutatnak, hogy míg a hagyományos irodalomtudomány egy antropocentrikus középleptékkal dolgozott, addig a számítógépes kutatók mindkét irányban „az irodalmi skála végleiteibe” hatoltak, és „a mikro és makro ezen mixe a digitális bölcsészet kézjegyévé vált, mely drámai hatást gyakorolt arra a léptékre, amelyen az irodalmi kutatások folynak”.<sup>40</sup> A szerzők azután a „bekezdés” (*paragraph*) képében egy újra-feltalált középleptéket kínálnak, hogy bemutassák, miként „aktiválnak a különböző léptékek különböző szerkezeti vonásokat”, s ennek megfelelően miként „függ össze közvetlenül a lépték a szöveg-funkciók differenciálódásával”.<sup>41</sup> A tanulmány furfangos (burkoltan túlhangsúlyozott) újrafeltalálását nyújtja Voltaire *Mikromegasz* című filozófiai novellájának is azzal, hogy a nyitófejezet címévé teszi. Mindezek teljes mértékben jogos kijelentések és gesztusok, ám a léptékváltás státusának megválaszolatlan kérdése akkor is velünk marad. Ha igaza volt Poe szövegének, és az elemzésnek túl kell haladnia a pusztán matematikai okoskodáson, akkor a léptékek tisztán kvantitatív logikája nem lesz elég. Léptékek nélkül persze semmire nem megyünk. Az olvasás csakis ezen vagy azon a konkrét léptéken mehet végbe. De az olvasás nem pusztán léptékek kérdése. Készen kell állnunk nem csupán léptéket váltani, de leszámolni is minden léptékkal, túllépni minden léptéken. Ha azt akarjuk, hogy az olvasás invenciózus legyen, hogy sosem látott találatokkal vagy találmányokkal álljon elő, melyek nem csupán váratlanok, de új utakat és távlatokat is nyitnak további színterekre, akkor azon kell lennünk, hogy az olvasás meghaladja a térképes játékot, a közelítés és távolítás, a lassítás és gyorsítás játékát, mely pusztán léptékekkel való játszadozás. Radikálisan eszkárolódnia kell: játéka nem merülhet ki e játék szélsőségig vitelében, hanem meg kell tanulnia túl is lépni ezen a játékon, minden léptéken túlra helyezkedni, és „eszkárolódnia”, ahogy minden igazi olvasás mindig is tette – ha egyáltalán létezik olyasmi, amit úgy hívhatunk: olvasás.

<sup>37</sup> Ha már a nem azonos ismétléseknél tartunk, ideje utalnom rá, hogy ez a mostani elemzés harmadik nekirugaszkodás részemről a szoros és a távoli olvasás kérdéskörének, s ennek kapcsán Poe novellájának. Lásd az alábbi, korábban készült, megjelenés alatt álló írásokat: „Szoros olvasás, távoli olvasás, sebességhatárok,” *Tiszatáj*, várható megjelenés: 2020. november (a téma hazai diskurzusára vonatkozólag lásd az említett tanulmány 12. lábjegyzetét); „Távoli pillantás a szoros olvasásra: iróniáról és terrorizmusról 1977 körül,” *Literatura* (2020): megjelenés alatt.

<sup>38</sup> Hancher, „Re: Search and Close Reading,” 128; Jin, „Problems,” 115.

<sup>39</sup> James F. English and Ted Underwood, „Shifting Scales: Between Literature and Social Science,” *Modern Language Quarterly* 77, 3. sz. (2016): 292, <https://doi.org/10.1215/00267929-3570612>.

<sup>40</sup> Mark Algee-Hewitt, Ryan Heuser and Franco Moretti, „On Paragraphs: Scale, Themes, and Narrative Form,” *Literary Lab: Pamphlet* 10. sz. (2015): 4 és 1, <https://litlab.stanford.edu/LiteraryLabPamphlet10.pdf>.

<sup>41</sup> Algee-Hewitt, Heuser and Moretti, „On Paragraphs,” 21.

## Escalated Reading: Stanford, Benjamin, Poe

In recent decades, the controversy over distant *versus* close reading has revolved around the question of scaling. Participants in the debate have either advocated distance (or speed), or have insisted on proximity (or slowness). On a meta-critical level, some have even argued for the need for any reading to be able to shift between, and thus combine, different scales. Very little has been said, however, about the limitations of scaling as such, and the irreducibility of reading to the logic of scales. Starting out from a few intricate formulations by some proponents of close and distant reading, this paper attempts to investigate the potentials and limitations of scaling, first by references to “Stanford” (the university as well as its founder), then by looking into Walter Benjamin’s treatment of film, and finally, though most importantly, by re-reading some passages in Poe’s detective story “The Purloined Letter.” These three points of reference (Stanford, Benjamin, Poe) seem analogous in the way they lay mutual emphasis on both serialization and segmentation, fast and slow motion, or distance and proximity. On a closer (or more distant?) look, however, Poe’s text goes even beyond such a scheme of scaling. It testifies to a logic of detection which surpasses mere zooming-in or zooming-out strategies, and points to a notion of reading that is “escalated” not simply because of its extraordinary range in terms of velocity or distance, but more radically because, although it still binds reading to specific scales, it also has an aspect that is utterly heterogeneous to any logic of scaling. The paper attempts to highlight this radically “escalated” (out-of-scale) aspect of reading.

Keywords:

close reading, distant reading, search, range, scale, escalation

**Fülöp Endre**

*Qulto-Monguz*

efulop@monguz.hu

## Szemantikus katalógus mint digitális bölcsészet

A könyvtárak és a digitális bölcsészet viszonya kapcsán az előbbi általában az utóbbi feltételeinek megteremtőjeként kerül említésre. A könyvtár gondolja azt a digitális ökoszisztémát, amelyben a digitális bölcsészeti projektek megvalósulhatnak. A könyvtár azonban közvetlenül és alkotó módon is részese lehet digitális bölcsészeti vállalkozásoknak, hiszen nemcsak a filológia, a textológia, a nyelvészet, az irodalom- vagy történettudomány válhat digitálissá, hanem a könyvek leíró felsorolásával foglalkozó tudomány, a könyvészet is. Dolgozatomban amellet érvelek, hogy a hagyományos könyvtári cédulakatalógus hálózati kiadásának elkészítése, azaz a szakterületi tudásrendszerként is működő szemantikus katalógus létrehozása voltaképpen a digitális bölcsészet értelmében vett digitális könyvészet.

Kulcsszavak:

digitális bölcsészet, digitális könyvészet, szemantikus katalógus, tudásgráf, közösségi tudomány, hálózati könyv



### 1. Bevezetés

Mindazt, amiről a következőkben szó lesz, az olvasók egy része talán magától értetődőnek fogja találni. Minden bizonnyal sokkal inkább ez a magyarázata annak, hogy a könyvtárak és a digitális bölcsészet viszonyának kapcsán oly kevés szó esik az alább olvasható összefüggésekről, nem pedig a gondolatok újszerűsége vagy az, hogy eddig elkerülte volna mindez a téma kutatóinak figyelmét. Akár így van azonban, akár úgy, tény, hogy valamilyen okból a következő oldalakon tárgyalt kapcsolódási pontok könyvtár és digitális bölcsészet között kevés figyelmet kapnak, s így méltatlanul háttérbe szorulnak más – nem kevésbé érdekes és releváns – szempontok mögött. Ez pedig azért nem szerencsés, mert a könyvtárak jövőbeli szerepéről való gondolkodásban ezeket az összefüggéseket is érdemes szem előtt tartani.

### 2. Digitális bölcsészet

Mielőtt azonban a könyvtár és a digitális bölcsészet közti kapcsolat elemzésébe fogunk, nem árt tisztázni, hogy pontosan mit is értünk digitális bölcsészet alatt. Patrick

Sahle, a téma elismert kutatója szerint „a digitális bölcsészet kutatási terület, sőt valószínűleg önálló tudományág is, de mindenképpen feladatkör”.<sup>1</sup> Ez a megfogalmazás jól kifejezi azt a bizonytalanságot, amely ezt a terminust körülengi. Új tudományágról van szó? Hagyományos diszciplínák által használt új módszerről? Vagy inkább új együttműködésről olyan területek között, amelyre korábban nem volt jellemző a közös gondolkodás, a közös munka? Fontos ez a kérdés egyáltalán? Bizonyos tekintetben igen. Annyi egyértelműnek látszik, hogy a digitális bölcsészet inter- vagy transzdiszciplináris terület, amelyhez legalább kétféleképpen közelíthetünk. Ha módszerként vagy tudományterületként tekintünk a dologra, akkor egy személyt, egy kutatót feltételezünk, valakit, aki digitális bölcsészetet űz. Ő a digitális bölcsész. Ha azonban inkább együttműködésként gondolunk rá, akkor ez már mindenképpen több személyt, több kutatót feltételez, akik mind a saját – hagyományos – tudományterületükön dolgoznak, digitális bölcsészetet ebben az értelemben, önmagában véve egyikük sem űz, a digitális bölcsészet az együttműködésük révén, az együttműködésük során születik meg. Ebben az esetben szigorúan véve digitális bölcsész nincs is, csak digitális bölcsészet.

A következőkben a digitális bölcsészet együttműködés-jellegét szeretném kiemelni, s az ebből születő szinergia fontosságát hangsúlyozni a digitális bölcsészeti projektek esetében, így az ELTE Digitális Bölcsészet Központjának bemutatkozó oldalán olvasható definícióra<sup>2</sup> fogok támaszkodni, és digitális bölcsészet alatt bölcsészettudományi kutatásokat és ezeket, valamint ezek eredményeinek disszeminációját támogató, kifejezetten erre a tudománycsoportra szakosodott számítástudományi és tudásmérnöki megoldások termékeny kollaborációjának elméleti kutatással is megtámogatott gyakorlatát fogom érteni.

### 3. Digitális könyvészet

Milyen szerepe lehet ebben az együttműködésben a könyvtárnak, a könyvtárosnak? Koltay Tibor szerint

több olyan témát azonosíthatunk, amely a könyvtártudományi kutatás mellett a digitális bölcsészet érdeklődésére is számíthat.

Ezek a következők:

- a dokumentumok tartalmi feltárása,
- a digitalizálás és a digitális dokumentumok (szövegek) megőrzése,
- a digitális könyvtárak építése,
- a publikációkhoz való nyílt hozzáférés,
- az információvisszakeresés,
- az eleve digitális dokumentumok,

<sup>1</sup> Patrick Shale, „Digital Humanities Als Beruf,” *Journal der Bayerischen Akademie der Wissenschaften* 1. sz. (2016): 78–83, 78.

<sup>2</sup> Az ELTE DH weboldalán egészen pontosan az alábbi meghatározás olvasható: „»Digital humanities« (DH) néven mára már világszerte elterjedt a bölcsészettudományok és a kifejezetten erre a tudomány-csoportra szakosodott informatika termékeny összekapcsolásának elméleti kutatással is megtámogatott gyakorlata.”, hozzáférés: 2020.09.15, <http://elte-dh.hu/hu/a-kozpontrol/>.

- a digitalizálás és a digitális megőrzés,
- a nyílt hozzáférés.<sup>3</sup>

Kokas Károly pedig arról ír, hogy a könyvtár a digitális ökoszisztéma gazdjaként kaphat szerepet.<sup>4</sup> Hozzátehetjük, hogy olyan digitális ökoszisztémának a gazdjaként, amely – egyebek mellett – digitális bölcsészetet is lehetővé tesz, tehát amelyben digitális bölcsészeti projektek megvalósulhatnak. Ez pedig tulajdonképpen egybeesik Koltay Tibor felsorolásával, hiszen az általa említett témák is inkább hozhatóak összefüggésbe a digitális bölcsészet feltételeinek megteremtésével, mint magával a szűkebb értelemben vett digitális bölcsészeti munkával.

A könyvtártudománynak, így a könyvtárosi munkának önmagában is van azonban bölcsészettudományi vonatkozása, lévén a könyvtártudomány – részben legalább – bölcsészettudomány. Így a könyvtár nem csupán közvetetten, a háttér megteremtőjeként, a szükséges feltételek biztosítójaként és gondozójaként kaphat szerepet a digitális bölcsészet területén, hanem közvetlenül, az ökoszisztéma használójaként, digitális bölcsészeti projektek alkotó tagjaként is. Erről hajlamosak vagyunk elfeledkezni, aminek minden bizonnyal az a legfőbb oka, hogy digitális bölcsészet alatt elsősorban digitális filológiát, textológiát, nyelvészetet, esetleg történelem- vagy irodalomtudományt értünk, digitális könyvészetet<sup>5</sup> sokkal ritkábban. Míg tehát mára megszoktuk, elfogadtuk, hogy a nyelvészet, a filológia, az irodalom- vagy a történettudomány együttműködhet a tudásmérnökséggel és a számítástudománnyal, s ebből digitális bölcsészet születhet, addig a könyvészet – értve ezalatt a „könyvek leíró felsorolásával foglalkozó tudomány[t]”<sup>6</sup> – vonatkozásában ugyanilyen együttműködésre általában nem gondolunk, vagy ha gondolunk is, nem digitális bölcsészetként.<sup>7</sup> Ennek persze a legfőbb oka valószínűleg az lehet, hogy a könyvészet területén a számítógép használata már régóta bevett gyakorlat. Elég a számítógépes katalógusokra, az OPAC-okra

<sup>3</sup> Koltay Tibor, „Gondolatok a digitális bölcsészet, a könyvtártudomány és a könyvtárak kapcsolatrendszeréről,” *Digitális Bölcsészet* 2 (2019): 3–13, 8–9, <https://doi.org/https://doi.org/10.31400/dh-hun.2019.2>.

<sup>4</sup> „Természetesen most ebben a kontextusban a gondolatmenet értelme az, hogy e folyamat végén (vagy valahol a folyamatban) megjelenik a – az általuk használt fogalom értelmében – digitális ökoszisztéma gazdája a könyvtár és a könyvtáros is, hogy az emlékezetmegtoldás eszközeit prezentálja, s arról megfelelőképpen tudósítson, segítve egy újfajta munkastílus, gondolkodás kialakítását.” Kokas Károly, *Kalauz a modern könyvtárak világába* (Budapest: Akadémiai Kiadó, 2020), 18, <https://doi.org/10.1556/9789634545736>.

<sup>5</sup> A „digitális könyvészet” keresőkérésre a legismertebb internetes keresőprogram 2020. szept. 16-án mindössze egyetlen találatot adott. Összehasonlításképpen: a „digitális filológia” kifejezésre 761 találatot kapunk, „digitális irodalomtudomány” kifejezésre pedig 407-et.

<sup>6</sup> *A magyar nyelv értelmező szótára*, főszerk. Bárczi Géza és Országh László (Budapest: Akadémiai Kiadó, 1959–1962). A továbbiakban könyvészet alatt ezt a tudományt értem.

<sup>7</sup> Akad persze kivétel is: Kokas Károly Szinnyei József (1830–1913) könyvészeti munkája kapcsán így fogalmaz: „Ha valaki ismeri az anyagot, s kicsit könyvtáros, lexikográfus szemmel tekint rá, világossá válik, hogy nagyrészt az történt, történhetett, hogy Szinnyei »kifordította« a meglévő hatalmas cédulaanyagát, s amit tematikusan már egyszer elrendezett, azt most – sok adattal kiegészítve persze – új sorrendbe, szerzői betűrend szerint »újrahasznosította«. Véleményem szerint az óriási adatmennyiség és a metódus együtt nyugodtan tekinthető a hazai digitális bölcsészet egyik, számítógép előtti előzményének.” Kokas Károly, „Digitális bölcsészet 2016,” in *MONOKgraphia: Tanulmányok Monok István 60. születésnapjára*, szerk. Nyerges Judit, Verók Attila és Zvara Edina (Budapest: Kossuth Kiadó, 2016), 405–412, 407.

gondolni. A számítógépesített könyvészet azonban önmagában nem jelent digitális bölcsészet értelmében vett digitális könyvészetet.<sup>8</sup> E különbségre – ti. a számítógépesített könyvészet és a digitális könyvészet közti különbségre – a dolgozat vége felé még visszatérek, jelen pillanatban elegendő talán annyit előrebocsátani, hogy nem tévedünk nagyot, ha egy hagyományos könyvtári OPAC és egy új generációs, szemantikus könyvtári katalógus közti különbséggént gondolunk rá.

Ez a párhuzam azt sugallhatja, hogy a szemantikus katalógusok építését digitális bölcsészeti projektnek tekintem. Ez a sugalmazás nem véletlen, a következő alfejezetekben éppen e tézis mellett szeretnék érvelni.

#### 4. Szemantikus katalógus

Mielőtt azonban erre rátérnénk, szükségesnek látok még egy rövid terminológiai kitérőt. A szemantikus katalógus kifejezésnek ugyanis a szakirodalomban két, egymással ugyan több tekintetben összefüggő, egymástól mégis jól megkülönböztethető, sőt bizonyos értelemben egymásnak némileg ellent is mondó (amennyiben ellentétes irányba mutató) jelentésével találkozhatunk.<sup>9</sup> Az egyik jelentés a *linked open data* („kapcsolt nyílt adat”) elmélete és gyakorlata felől közelít ehhez a fogalomhoz, s akkor tekinti a könyvtári katalógust szemantikusként, ha az képes külső rendszerek, mindenekelőtt internetes keresőprogramok számára jól hasznosítható, könnyen hozzáférhető és pontosan értelmezhető információforrássul szolgálni, azaz ha szervesen és szorosan integrálódik a szemantikus webbe.

A következőkben azonban nem ebben az értelemben fogom a szemantikus katalógus kifejezést használni, az értelmezési keretet nem a szemantikus web fogja jelenteni. A szemantikus katalógus összetétel a jelzőt tehát nem a szemantikus webtől örökli. De ha nem onnan származik, akkor mit kell alatta érteni? Az alábbi meghatározás, úgy vélem, jó kiindulópontot jelent ennek a kérdésnek a megválaszolásához, hiszen tartalmazza a legfontosabb mozzanatokat:

A szemantikus technológiát olyan szoftvertechnológiaként definiáljuk, amely az információk jelentésének és a köztük fennálló kapcsolatoknak a megismerését és feldolgozását futási időben lehetővé teszi. Ahhoz, hogy egy szemantikus technológia valóban működőképes legyen egy rendszerben, rendelkezésre kell hogy álljon a világ egy részének ismeretmodellje.<sup>10</sup>

E meghatározás alapján akkor illethetünk egy könyvtári katalógust a szemantikus jelzővel, ha benne rendelkezésre áll a könyvészeti tudásterület ismeretmodellje. Másként

<sup>8</sup> A digitális könyvészet kifejezés nem tekinthető a tudományos közösség által széles körben használt, bevett terminusnak, így természetesen pontos jelentéséről sem alakulhatott ki konszenzus. Jelen dolgozat egyik célja, hogy ehhez a jelentésadási folyamathoz szerény mértékben hozzájárulva segítse e kifejezés elterjedését és használatba vételét.

<sup>9</sup> A két jelentésről részletesebben lásd Fülöp Endre, „A szemantikus háló két fogalma, a katalógusok új generációja és a könyvtárak szerepe,” *Tudományos és Műszaki Tájékoztatás* 65, 7–8. sz. (2018): 401–408.

<sup>10</sup> Polikoff, Irene and Dean Allemang, *Semantic technology: TopQuadrant Technology Briefing*, <https://lists.oasis-open.org/archives/regrep-semantic/200402/pdf00000.pdf>. Idézi Munk Sándor, „Szemantika az informatikában,” *Hadmérnök* 9, 2 sz. (2014): 311–331, 318.

fogalmazva: ismeri és kezeli ennek a tudásterületnek valamennyi fontos fogalmát, valamint ezek minden lehetséges szemantikai kapcsolatát. A számítógépesített katalógusok, a jelenleg használt OPAC-ok esetében ez a kritérium nem teljesül, hiszen adatmodelljünkben megjelenő entitások köre a könyvészet fogalmainak csupán szűk részhalmazára terjed ki. Kis túlzással<sup>11</sup> azt mondhatjuk, hogy a számítógépes katalógusok adatmodelljében csupán két fogalom – a kiadás és a példány – jelenik meg teljes értékű entitásként. Ahhoz, hogy egy könyvtári katalógust a fenti értelemben szemantikusnak nevezhessünk, a tudásterület szempontjából releváns, ám a jelenlegi OPAC-ok által nem kezelt fogalmakat be kell emelni a katalógus tudásmodelljébe – a fogalmak közti szemantikai kapcsolatok leírásával együtt. Hogy milyen fogalmakról és milyen kapcsolatokról van szó, erre a következő alfejezetekben láthatunk néhány példát.

Mindez szorosan összefügg a szemantikus katalógus tudásrendszer-jellegével is. A hagyományos katalógusok célja végső soron a dokumentumszolgáltatás. Az OPAC-ok elsődleges feladatának azt tartjuk, hogy a felhasználókat hatékonyan, egyszerűen és gyorsan elvezessék az őket érdeklő dokumentumokhoz – legyen az akár nyomtatott, akár elektronikus dokumentum. A szemantikus katalógus a tudásterület teljesebb és pontosabb leképzése révén alkalmassá válik arra, hogy a dokumentumokhoz kapcsolódó felhasználói kérdések minden eddiginél szélesebb körére választ tudjon nyújtani. Tehát arra, hogy a hagyományos katalógusfunkciók mellett tudásszolgáltatást is kínáljon a felhasználóknak. Erre szintén látunk példát a következőkben.

Összefoglalva tehát: szemantikus katalógus alatt a továbbiakban olyan szakterületi tudásalapú rendszert értek, amely a hagyományos könyvtári katalógusokban megszokott fogalmakon túl, a könyvészeti tudásterület valamennyi releváns fogalmát leképezi, érti, kezeli, ezek egymáshoz való (szemantikai) kapcsolatát pontosan leírja, így a dokumentumszolgáltatás mellett tudásszervezésre és -szolgáltatásra is alkalmas. A dolgozat tézise – mely szerint szemantikus katalógus építését joggal tekinthetjük digitális bölcsészetnek – ilyen tulajdonságú könyvtári katalógusra vonatkozik.

## 5. Tudásgráf<sup>12</sup>

Nézzünk konkrét példát arra, hogyan építhető ilyen szemantikus katalógus! Fontos megjegyezni azonban, hogy mindabban, amiről a következő alfejezetekben szó lesz, a folyamaton és az együttműködésen van a hangsúly, és nem az eredményen. A választott szövegek és az ábrák csupán illusztráció gyanánt szolgálnak, azt hivatottak megmutatni, hogy a gráfalapú tudásreprezentáció és szemantikus katalógus milyen megközelítést jelent és milyen lehetőségeket rejt magában. Ennek a megközelítésnek és folyamatnak az egyik nagy erénye éppen az, hogy szélesebb körű együttműködésre épülve a javítás, pontosítás lehetősége mindig nyitva áll a résztvevők előtt. Az ábrákon megjelenő tudásgráf és szemantikus katalógus így nem egy folyamat eredményeként értelmezendő, hanem sokkal inkább kiindulópontként. Az esetleges hiányosságok,

<sup>11</sup> Az analitikus és az authority-rekordok révén olykor megjelennek más entitások is.

<sup>12</sup> A tudásgráf helyett írhatnánk szemantikai hálót is. Ez utóbbi terminus azonban túlságosan terhelt a *szemantikus web*bel kapcsolatos asszociációkkal, s mivel ezeket lehetőség szerint kerülni szeretném, a továbbiakban elsősorban az előbbi kifejezést fogom használni.

pontatlanságok így csak annyiban érdekesek, amennyiben ezek javításának, pontosításának lehetőségére vagy lehetetlenségére irányítják a figyelmet.

Az 1955-ös, hatkötetes Shakespeare-összkiadás<sup>13</sup> minden drámához tartalmaz egy-egy rövid bevezetőt, amelyet az adott mű avatott ismerője, kutatója írt. Ezek a bevezető tanulmányok minden esetben kitérnek a drámák magyar nyelvű fordításaira is. A *Lear király* kapcsán például Kéry László irodalomtörténész, akadémikus tollából – többek között – ezt olvashatjuk:

A Nemzeti Színház egészen 1865-ig Jakab István és Vajda Péter fordítását játszotta, pedig 1856-ban már megjelent Vörösmarty fordítása. Vörösmartynak erről a vállalkozásáról Petőfi már 1848-ban hírt ad egy Aranyhoz írt levelében. A költő a munka zömével valószínűleg 1852-53-ban készült el, de még 1855-ben is dolgozott rajta. [...] A fordításon a Kisfaludy Társaság sorozatában való megjelenés (1865) előtt Arany igazított. Vörösmartynak ezt a munkáját általában nem tartják olyan egyenletesnek, mint Caesar-fordítását, de szépségekben, remek költői megoldásokban ez is bővelkedik. Századunkban Kosztolányi Dezső, és újabban Füst Milán fordította le a *Lear királyt*.<sup>14</sup>

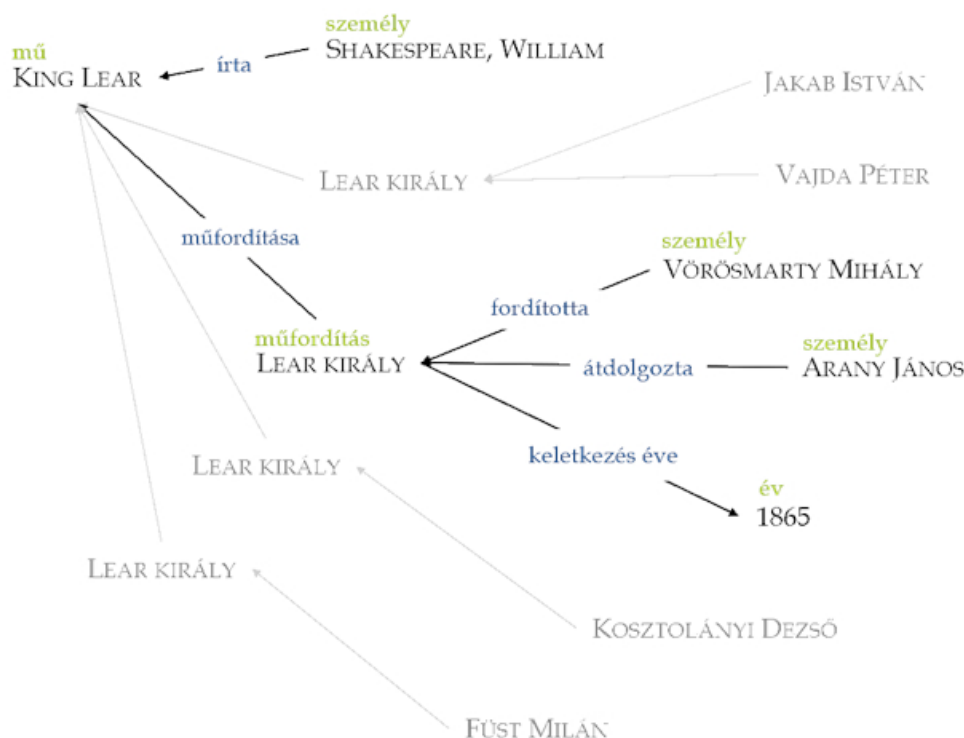
Ebben a bekezdésben nagyrészt olyan információkat olvashatunk, amelyek kiválóan ábrázolhatóak tudásgráfban is. Az első ábrán Kéry László által közölt tudáselemek gráfos reprezentációját láthatjuk. A jobb áttekinthetőség érdekében az entitások és kapcsolatok osztályozását nem minden csomópont és él esetében végeztem el. A szűrővel jelölt, az ábrán osztályba nem sorolt csomópontok és élek értelmezése a feketével jelöltek alapján azonban, úgy vélem, mindenki számára könnyedén elvégezhető.

Szükségtelen hosszan érvelni amellest, hogy milyen előnyökkel jár, ha az információkat tudásgráfban ábrázoljuk, és nem csak folyó szöveg formájában állnak rendelkezésre. Elegendő arra utalni, hogy ezzel a tudás számítógépes rendszerek jóval szélesebb köre számára válik feldolgozhatóvá, érthetővé. A digitális bölcsészet részben éppen annak a felismerésnek köszönheti létét, hogy az ilyen tudásreprezentációs módszerek és eszközök a bölcsészettudományok területén is kiváló szolgálatot tehetnek.

<sup>13</sup> William Shakespeare, *Shakespeare összes drámái I–IV*. (Budapest: Új Magyar Könyvkiadó, 1955).

<sup>14</sup> Kéry László, „Jegyzetek a *Lear királyhoz*,” in William Shakespeare, *Shakespeare összes drámái III. kötet: Tragédiák* (Budapest: Új Magyar Könyvkiadó, 1955), 639–677, 648.





1. ábra. Tudásgráf a *Lear király* magyar fordításairól  
(Kéry László tanulmánya alapján)

A fenti tudásgráf – meglehetősen egyszerűsége folytán – azt az érzést keltheti, hogy ilyenfajta hálót, illetve a háló alapjául szolgáló ontológiát megtervezni egyszerű feladat. Ez azonban csalóka benyomás. Aki tervezett már valaha ontológiát, biztosan meg tudja erősíteni, hogy milyen hamar eljut az ember olyan kérdésekig, amelyekre jó választ adni és helyes döntést hozni megfelelő szakértelem és tapasztalat hiányában nagyon nehéz. Két példát hadd említsek erre! Az egyik a mű keletkezési időpontjának kérdése. Közismert, hogy Shakespeare művei esetében rendkívül nehéz egyértelmű választ adni arra, hogy mit kell egy-egy mű keletkezési évének tekinteni. A kvartó kiadás időpontját? Vagy inkább a fólió kiadását? Esetleg – ha ismert és korábbi a kiadásokénál – a bemutató előadás időpontját?<sup>15</sup> Vajon szükség van-e arra, hogy a tudásgráfban leképezzük ezt a problémát? Ontológiatervezés során ez a kérdés egészen pontosan abban a formában merül fel, hogy meg kell-e jelenniük a tudásmodellben az olyan fogalmaknak, mint a kvartó kiadás vagy a fólió kiadás? Hasonló a helyzet a különböző fordítások esetében is. Varró Dániel *Lear*-fordításának viszonya az eredeti műhöz kétségkívül más, mint például Nádasdy Ádámé. De azt jelenti-e ez, hogy a tudásgráf

<sup>15</sup> A *Lear király* esetében például ezt olvashatjuk a kísérőtanulmányban: „Nyomtatásban először 1608-ban jelent meg, de már 1607-ben szerepelt a kiadásra szánt könyvek listáján Shakespeare nevével és azzal a megjegyzéssel, hogy a király előtt »a múlt karácsonykor« eljátszották. Valószínűleg 1605-ben vagy 1606-ban keletkezett. [...] Az 1608-as kvartó-szöveg – amelynek egyik változatából 1619-ben újabb negyedréti kiadás készült – színészi diktálás, esetleg gyorsírási feljegyzések alapján keletkezett és háromszáz sorral többet tartalmaz, mint az első fólió-kiadás (1623).” Kéry, „Jegyzetek a *Lear király*hoz,” 639.

alapjául szolgáló ontológiában is meg kell különböztetni a kétféle relációt, és olyan fogalmakat is be kell vezetni, mint a „tradaptáció”? Ezekre a kérdésekre egyértelműen helyes válasz természetesen nincs, döntést azonban hozni kell. A digitális könyvészet tudásmodelljének tervezése során mi azt az elvet tartottuk szem előtt, hogy azokra a fogalmakra és relációkra támaszkodunk, melyeket a digitális könyvészetben együttműködő szakemberek mindegyike napi szinten használ és pontosan ért. A műfordítást ilyen fogalomnak tekintettük, a tradaptációt vagy a kvartókiadást nem. Így az előbbi megjelenik az ontológiában, az utóbbiak nem. Ez természetesen némileg önkényes döntésnek tűnhet, hozzátehetjük azonban ehhez rögtön azt is, hogy az ontológiaalapú tudásreprezentáció előnyét éppen az jelenti a hagyományos, rekordokra építő relációs adatmodellezéssel szemben, hogy a leképezett tudásterület sokkal rugalmasabban bővíthető, új fogalmak, új relációk beemelése jóval könnyebben elvégezhető. Így annak sincs igazi akadálya, hogy a későbbiekben most kihagyott fogalmak bekerüljenek az ontológia tudásmodelljébe. Erre is hozok példát a későbbiekben.

Az ontológiatervezés azonban nem csak a fentiekhez hasonló kérdések miatt igényel speciális tudást és képzettséget. Voltaképpen tudásreprezentációról, adatmodellezésről van szó, ezért is érdemes a digitális bölcsészet interdiszciplinaritása kapcsán az együttműködés fontosságát hangsúlyozni, és ezért szerencsés, ha a digitális bölcsészeti projekteken aktív szerepet vállalnak tudásmérnökök is.

Visszatérve a fenti példához: azt hiszem, amellet sem kell hosszasan érvelni, hogy a tudásgráfban reprezentált információk a könyvtári katalógus szempontjából is relevánsak. Az a tudás, hogy a *Lear királynak* hány magyar nyelvű fordítása van, ki fordította le és mikor, a könyvtárhasználók széles rétege számára lehet érdekes és fontos. Helye van tehát a katalógusban. Illetve helye lenne, mert jelenleg nincs meg benne. A hagyományos OPAC-okban ennek a tudásnak csupán egy kis része található meg, s az is nagyon nehezen nyerhető ki belőle. A fordító személyéről szóló adat a kiadásrekordokban még gyakran rendelkezésre áll, ám hosszú találati listákat kell végigböngésznie annak, aki a különböző műfordításokról teljes, átfogó képet akar nyerni. Sőt gyakran még ez sem elegendő, hiszen a több drámát tartalmazó köteteknél a fordítók személye és a fordított művek címe nincs egyértelműen összekapcsolva, egymáshoz rendelve (azaz nem azonosítható, hogy mely drámát ki fordította). Ezek a hiányosságok nyilván azzal függenek össze, hogy a könyvtári katalógusokkal szemben nem támasztunk tudásszolgáltatással, tudásrendszerrel kapcsolatos igényeket, a bibliográfiai információk rögzítésével és kezelésével kapcsolatos döntésekben szinte kizárólag a visszakereshetőség szempontjai érvényesültek.

Mindebből az következik viszont, hogy az irodalomtörténeti bevezetőből, irodalomtudástól származó és tudásmérnök által tudásgráfban reprezentált információval a könyvtárosoknak is lehet teendőjük. Ők rendelkeznek ugyanis azzal az eszköz-készlettel – értve ezalatt könyvészeti szakértelmet, bibliográfiai adatbázisokat és új generációs integrált könyvtári szoftvereket –, amely ahhoz szükséges, hogy a tudásgráfot lefelé tovább lehessen építeni – azaz, hogy az egyes műfordításokat jelölő csomópontokhoz konkrét kiadásokat reprezentáló csomópontokat lehessen kötni a tudásgráfban („tartalmazza” jelentésű éllel). Az az információ ugyanis, hogy a *Lear királynak* több fordítása van, úgy válik széles körben jól használható tudássá, ha

kiegészül azzal, hogy melyik fordítást hol – mely kiadó mely köteteiben vagy milyen digitális tartalomszolgáltatás keretében milyen weboldalon – találjuk meg.

Ezzel a könyvtárosi hozzájárulással válik a tudásgráf katalógussá, illetve – másfelől nézve – a katalógus szemantikussá.

**Szerző:** Shakespeare, William (1564 – 1616) költő, drámaíró  
**Cím:** KING LEAR / LEAR KIRÁLY  
**Műfaj:** mű / irodalmi mű / dráma  
**Nyelv:** angol  
**Első kiadás:** London, 1608  
**Tárgyszavak:** tragédia

**Kiadások** (eredeti szöveg) 6 kiadás  
**Műfordítások** 3 nyelv, 13 műfordító, 35 kiadás

- ☐ magyar, műfordító: **Sófai József** (1784) 1 kiadás ▲  
☐ Budapest Recli 2016 ➔
- ☐ magyar, műfordító: **Jakab István, Vajda Péter** (1838)
- ☐ magyar, műfordító: **Vörösmarty Mihály** (1856) 23 kiadás ▼
- ☐ magyar, műfordító: **Zsigány Árpád** (1899)
- ☐ magyar, műfordító: **Kosztolányi Dezső** (1936) 3 kiadás ▼
- ☐ magyar, műfordító: **Füst Milán** (1955) 4 kiadás ▲
  - ☐ Budapest Új Magyar Könyvtár 1955 ➔
  - ☐ Budapest Európa 1957 ➔
  - ☐ Budapest Magvető 1966 ➔
  - ☐ Budapest Magvető 1979 ➔
- ☐ magyar, műfordító: **Mészöly Dezső** (1986) 5 kiadás ▲
  - ☐ Budapest Európa 1999 ➔
  - ☐ Budapest Európa 2001 ➔
  - ☐ Budapest Nemzeti Színház 2008 ➔
  - ☐ Budapest Európa 2014 ➔
  - ☐ Budapest Európa 2018 ➔
- ☐ magyar, műfordító: **János György** (2002) 1 kiadás ▲
  - ☐ Marosvásárhely Mentor 2002 ➔
- ☐ magyar, műfordító: **Varró Dániel** (2009) 1 kiadás ▲
  - ☐ Budapest Nemzeti Színház 2010 ➔
- ☐ magyar, műfordító: **Nádassdy Ádám** (2010) 3 kiadás ▲
  - ☐ Budapest Magvető 2007 ➔
  - ☐ Budapest Magvető 2012 ➔
  - ☐ Budapest Magvető 2017 ➔
- ☐ német, műfordító: **Rapp, Moritz** (1842) 1 kiadás ▲

2. ábra. A *Lear király* rekordoldala egy szemantikus katalógusban<sup>16</sup>

A második ábrán azt láthatjuk, hogyan épül fel Shakespeare *Lear király*ának rekordoldala egy szemantikus katalógusban. A hogyan alatt itt elsősorban struktúra és nem adattartalom értek. Az utóbbi tekintetében lehetnek rajta pontatlanságok, hiányosságok, ezek természetesen javítandók, bővítendőek. A hangsúly azonban – ahogy erről korábban is esett már szó – nem az eredményen, hanem a folyamaton van. A szemantikus tudásrepresentáció révén olyan fogalmakkal és olyan relációkkal dolgozhatunk a katalógus építése során, amelyeket korábban nem használhattunk. A fenti ábra azt szemlélteti, hogy ezekből a fogalmakból és relációkból épített gráf hogyan jelenhet meg felhasználói felületen. Ne befejezett tudásgráfot lássunk tehát az ábrában, hanem olyan keretet, amely tudásgráf építésének alapjául szolgálhat, tartalommal megtöltve pedig közelebb vihet egy pontosabb, teljesebb tudásgráfhoz.

Az irodalomtörténész, a tudásmérnök és a könyvtáros együttműködése révén az olvasó gyorsan és könnyen – ahogy a tudásgráf épül, bővül – egyre pontosabb, hitelesebb áttekintést kap a dráma magyar műfordításairól úgy, hogy mindez szervesen

<sup>16</sup> A szemantikus katalógusról készült ábrák a Debreceni Egyetem Egyetemi és Nemzeti Könyvtár és a Monguz Kft. közös K+F projektjének részeként elkészült pilot FRBR-katalógus <https://opac3.frb.r.monguz.hu/> (hozzáférés: 2020.09.15.) címen elérhető publikus felületeit felhasználva készültek.

integrálódik a jól ismert könyvtári katalógus információszeretébe. Ez a műfordításokról szóló tudás kutatók, szakértők számára talán jelenleg is hozzáférhető különböző forrásokból – tanulmányokból, monográfiákból, lexikonokból vagy jegyzetektől, bevezetőkből –, ám nekik is, és főleg a területen kevésbé jártas diákok, hallgatók, érdeklődők, kezdő kutatók számára lehet különösen fontos és előnyös, ha a tudásháló építése révén könnyen elérhető, naprakész, hiteles, pontos, áttekinthető, jól használható forrás áll elő. S talán egyetérthetünk abban is, hogy ennek a könyvekről szóló, megbízható, jól szervezett, könnyen áttekinthető tudásnak a könyvtári katalógusoknál jobb, megfelelőbb helyet nehezen találhatunk.

A szemantikus katalógus ilyen formában irodalomtörténeti kutatások eredményeinek disszeminációját támogathatja, ezen keresztül pedig segítségül, inspirációul, kiindulópontul szolgálhat elsősorban diákok, kezdő kutatók, doktoranduszok, egyetemi hallgatók számára ugyan, de talán nem csak nekik. Az történt tehát, hogy a tudásmérnökség és – új generációs könyvtári szoftverek megalkotása révén – az informatika együttműködve az irodalomtörténettel és a könyvészzel a bölcsészettudomány szolgálatába állt. Ez pedig pontosan megfelel a digitális bölcsészet fentiekben idézett definíciójának. A digitális bölcsészet olyan területéről van tehát szó, amelyben a könyvtárosoknak – könyvészeti szakértelmük révén – aktív, közvetlen, alkotó szerep jut.

## 6. „Közösségi tudomány”

Ha a digitális bölcsészet kapcsán az együttműködés mozzanatát hangsúlyozzuk, nem hagyhatjuk említés nélkül a közösségi tudásépítés különböző formáit, a *crowdsourcing*-ot és a *citizen science*-t, azaz „közösségi tudomány”<sup>17</sup> sem. Mindkét esetben külső szereplők, jellemzően nagyobb online közösség segítségét veszik igénybe egy adott feladat elvégzésére, tartalom létrehozására vagy ötletek gyűjtésére. A különbséget az jelenti, hogy miközben *crowdsourcing*nak tekinthető minden olyan feladatmegosztás, amelyben intézményen kívüli felhasználók szerepet kapnak – függetlenül a feladat jellegétől és a résztvevők képzettségétől, valamint ismereteiről a projekt egészét illetően –, addig a *citizen science* esetében már tudományos jellegű és tudományos értékű közreműködésről beszélhetünk.

A résztvevő állampolgárok (innen a *citizen*) érdeklődést mutatnak a kutatási terület iránt, intellektuális erőforrásaikat bocsátják a kutatások rendelkezésére, valamint a kutatások vezetői is képzett szakemberek, tudományos kutatók. Emellett a civilek által elvégzett munka új ismeretanyagot, kutatható adatokat eredményezhet. Ezzel szemben előfordulhat, hogy egy *crowdsourcing* projekt résztvevője nincs is tudatában annak, hogy pontosan milyen eredményt hoz majd a közreműködése, nem feltétlenül hajtja a közös kutatási terület iránti érdeklődés.<sup>18</sup>

<sup>17</sup> Bár az alfejezet címeként a *citizen science* magyar fordítását használtam a magyar kifejezés terjesztése érdekében, a szövegben a könnyebb érthetőséget szem előtt tartva meghagytam a magyar szakirodalomban egyelőre még elterjedtebb angol terminust.

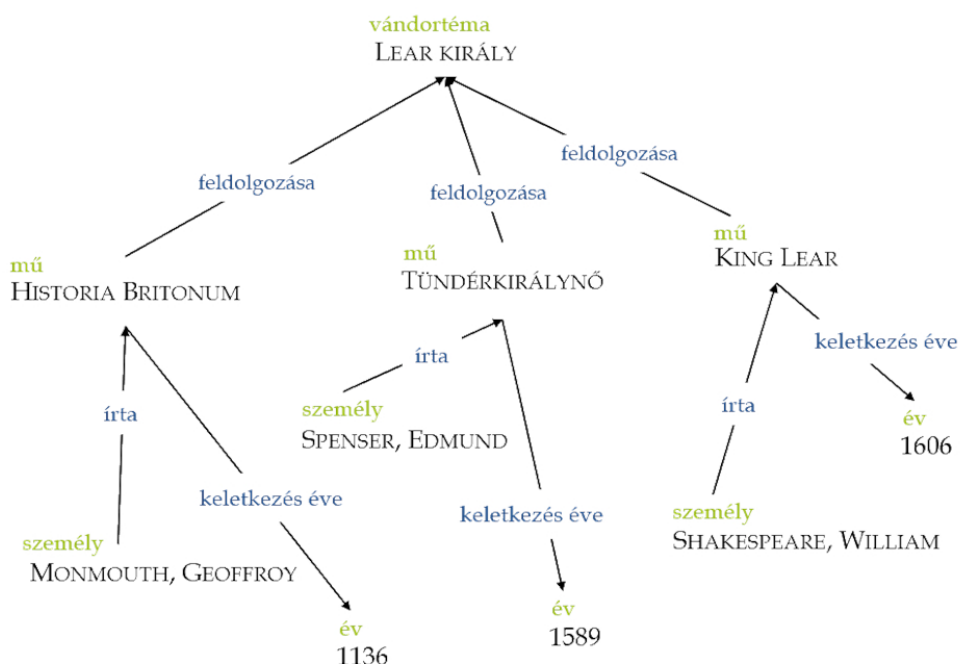
<sup>18</sup> Maczelka Árpád, „Az Országos Széchényi Könyvtár szerepe a Civic Epistemologies projektben,” *Tudományos és Műszaki Tájékoztatás* 62, 5. sz. (2015): 181–186, 181.

Az, hogy a *crowdsourcing* jelentősége a digitális bölcsészetben óriási, nem szorul igazolásra.<sup>19</sup> A következőkben egy példán keresztül azt szeretném bemutatni, hogy a szemantikus katalógus építése során a *citizen science* is szerepet kaphat.

Ismét Kéry László Lear királyhoz írt kísérőtanulmányából idézek:

Az öreg királynak és három lányának története nyilvánvalóan mondai termék. Az események idejét jóval a római hódítás előtti Britanniában az időszámításunk előtti VII. évszázadban kell elképzelnünk. A monda első irodalmi változatát Geoffroy of Monmouth *Historia Britonum* című latin nyelvű krónikájában találjuk, a XII. században. Ezek után főleg verses elbeszélések formájában tűnik föl. Benne van a középkor híres gyűjteményében, a *Gesta Romanorum*-ban, reneszánsz kori népszerűségéről pedig Az előljárók tükre és Spencer Tündérkirálynője tanúskodik.<sup>20</sup>

Az irodalomtudós által velünk megosztott ismeretek nagy része itt is jól megragadható tudásgráf formájában, ahogy ezt a második ábrán láthatjuk.



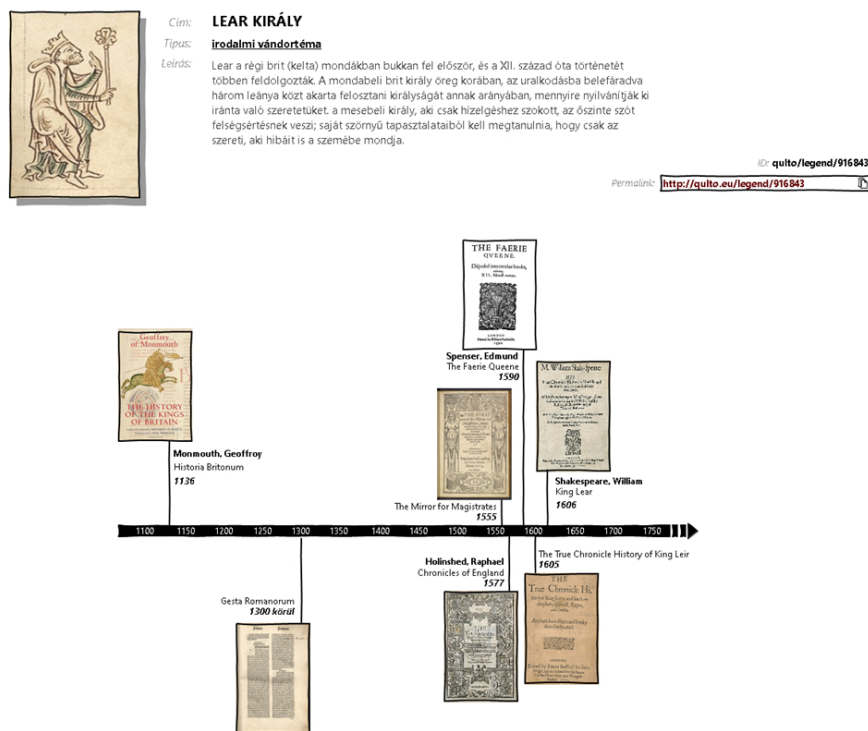
3. ábra. Tudásgráf Lear király mondájáról  
(Kéry László tanulmánya alapján)

A tudásgráfon ábrázolt ismeretek ráadásul közvetlen, szerves kapcsolatban állnak a könyvtári katalógusokban tárolt és kezelt bibliográfiai információkkal, így ez a tudás nem lenne idegen test egy a könyvtári katalógusban.

<sup>19</sup> Lásd például Melissa Terras, „Crowdsourcing in the Digital Humanities,” in *A New Companion to Digital Humanities*, eds. Susan Schreibman, Ray Siemens and John Unsworth (Chichester; Malden, MA: John Wiley & Sons, Ltd., 2016), 420–438, <https://doi.org/10.1002/9781118680605.ch29>.

<sup>20</sup> Kéry László, „Jegyzetek a *Lear király*hoz,” 639.

Ha a tudásmérnök megfelelő módon bővíti a tudásháló alapját képező ontológiát (az „irodalmi vándortéma” fogalommal és az arról szóló állításokat definiáló predikátumokkal), az informatikus pedig jól használható *citizen science* megoldást épít a katalógusba, akkor nincs semmi akadálya annak, hogy ez a *Lear király*hoz írt bevezetőben folyószöveggént olvasható tudás a szemantikus katalógus szerves részévé váljék, s a Shakespeare-dráma alapjául szolgáló irodalmi vándortéma is bekerüljön a katalógusba oly módon, hogy a vándortémát és annak feldolgozásait az érdeklődő és hozzáértő felhasználók a tudásgráfban konkrét művekhez és azon keresztül kötetekhez, példányokhoz kapcsolhatják.



4. ábra. A *Lear király* mondájának rekordoldala egy szemantikus katalógusban

Az irodalomtörténeti tudás, az irodalomtörténeti kutatás eredménye – tudásgráfban ábrázolva és infografika formájában bemutatva – a szemantikus katalógus részeként hiteles, pontos és könnyen befogadható, könnyen feldolgozható, könnyen hasznosítható tudássá válhat akár sokkal szélesebb kör számára is, mint jelenleg, amikor csak monográfiákban, tanulmányokban férhető hozzá a szakértő közönség számára. Ebben a formában és keretben inspirációul szolgálhat például szemináriumi dolgozatokhoz, iskolai kiselőadásokhoz, új kutatási témákhoz a könyvtárhasználó diákok, hallgatók számára.

Ismét olyan megoldással van tehát dolgunk, amelyben a bölcsészettudományi kutatás eredményének disszeminációját, hozzáférhetőségét, felhasználhatóságát növeli azáltal, hogy a bölcsészettudomány, a tudásmérnökség, az informatika és a könyvészet együttműködött. Tehát digitális bölcsészeti munka révén. Jóllehet ebben az esetben a

tudásbővítés maga elsősorban *citizen science* révén történt, a validáció, az ellenőrzés és különösen a hitelesítés feladata aktív szerepet ró a könyvtárosokra is.

## 7. Hálózati könyv

A digitális bölcsészet kulcsfogalmai közé tartozik a hálózati könyv (*networked book*) vagy hálózati kiadás is. A digitális bölcsészeti projektek igen jelentős része hálózati könyv vagy hálózati kiadás létrehozására irányul.

De mit takar ez a fogalom? A hálózati könyvre a könyv új formájaként, a könyv evolúciójának következő lépcsőfokaként szoktak hivatkozni. „Nyitott könyv, amelyet hálózati környezetben történő írásra, szerkesztésre és olvasásra terveztek.”<sup>21</sup> Közismert példaként a Wikipédiára utalhatunk. Négy fő, konstitutív jellemzőjét szokás kiemelni. Első, hogy strukturálisan szemcsés, azaz viszonylag kis darabokból, szövegekből, tudáselemekből áll össze. Második fontos jellemzője a hálózatiság (amelyről az új forma a nevét is kapta), ez a struktúra szemcséinek sokrétű összekapcsolására utal, illetve a kapcsolatok, linkek fontos szerepére. A harmadik és negyedik tulajdonsága pedig a nyitottság és a közösségi jelleg. A nyitottság itt elsősorban befejezetlenségre utal, arra, hogy szemben a nyomtatott könyvekkel itt semmi sem végleges, a hálózati könyv mindig folytatható, bővíthető, javítható, módosítható. A közösségi jelleg alatt pedig azt kell érteni, hogy a könyv kollektív alkotás, a hálózati könyvet közösség állítja össze. A szerzőség fogalma is jelentős mértékben átalakul tehát a hagyományos könyvhöz képest.

Ha az új könyv tulajdonságait nézzük, és a különbségekre a hálózati könyv újszerűsége felől pillantunk, akkor egyetérthetünk Horváth Iván megállapításával, mely szerint papírkönyv és e-könyv között olyan rettenetesen nagy különbség azért nincsen.<sup>22</sup> Az, hogy a hagyományos könyveket nem papíron, hanem képernyőn olvassuk, s hogy ez néhány hasznos új lehetőséget kínál (pl. keresésre, másolásra, a szöveg nagyítására stb.) valóban apróságnak, jelentéktelen különbségnek tűnik ahhoz képest, amit a hálózati könyv nyújt a korábbi formákkal összevetve. Az e-könyvek esetében valóban ugyanolyan módon előállított, ugyanolyan struktúrában közvetített tartalmakat fogyasztunk, mint a hagyományos papírkönyvekben. Ebben a tekintetben csak a hálózati könyvek megjelenésével állt be fordulat.

Hogyan kapcsolódik mindez a digitális könyvészethez és a szemantikus katalógushoz? Meglátásom szerint a könyvtári katalógusok evolúciója sok tekintetben analóg a könyvek fejlődésével. A hagyományos, papíralapú cédulakatalógusok helyett első körben megjelentek a számítógépes katalógusok, az OPAC-ok. De ha felidézzük egy átlagos OPAC találati listáját – például a „Lear király” keresőkérdésre (lásd 5. ábra) –, akkor azt mondhatjuk, hogy rettenetesen nagy különbség azért itt sincs. Persze – ahogy az e-könyv a hagyományos papírkönyvvel szemben – az OPAC is nyújt olyan lehetőségeket, amelyek a cédulakatalógusban nem álltak rendelkezésünkre: sokkal jobbak például a keresési, szűrési, exportálási lehetőségek. A tartalom tekintetében azonban – éppúgy, ahogy a papírkönyv és az e-könyv esetében – gyakorlatilag azonos

<sup>21</sup> Wikipédia, „Hálózati könyv – Networked book,” hozzáférés: 2020.09.21, [https://hu.qwe.wiki/wiki/Networked\\_book](https://hu.qwe.wiki/wiki/Networked_book).

<sup>22</sup> Horváth Iván, *Gépeskönyv* (Budapest: Balassi Kiadó, 2006).



a kettő. A hagyományos OPAC-okban voltaképpen a régi cédulák digitális változatai között böngészünk, ezeket lapozgatjuk, ezeket nézegetjük.

24.	<input type="checkbox"/>	Shakespeare, William (1564-1616)	<a href="#">Lear király /</a>	2017	Könyv	<a href="#">Részletek</a>
25.	<input type="checkbox"/>	Shakespeare, William (1564-1616)	<a href="#">Lear király /</a>	2017	Könyv része	<a href="#">Részletek</a>
26.	<input type="checkbox"/>	Shakespeare, William (1564-1616)	Négy dráma /	cop. 2017	Könyv	<a href="#">Részletek</a>
27.	<input type="checkbox"/>	Shakespeare, William (1564-1616)	<a href="#">Lear király /</a>	2014	Könyv része	<a href="#">Részletek</a>
28.	<input type="checkbox"/>	Shakespeare, William (1564-1616)	Öt dráma /	2014	Könyv	<a href="#">Részletek</a>
29.	<input type="checkbox"/>	Shakespeare, William (1564-1616)	<a href="#">Lear király</a> : tragédia öt felvonásban /	1966	Könyv része	<a href="#">Részletek</a>
30.	<input type="checkbox"/>	Shakespeare, William (1564-1616)	<a href="#">Lear király</a> = King Lear /	2012	Könyv része	<a href="#">Részletek</a>
31.	<input type="checkbox"/>	Shakespeare, William (1564-1616)	Három dráma /	2012	Könyv	<a href="#">Részletek</a>
32.	<input type="checkbox"/>	Shakespeare, William (1564-1616)	<a href="#">Lear Király</a> : tragédia öt felvonásban /	1979	Könyv része	<a href="#">Részletek</a>
33.	<input type="checkbox"/>	Shakespeare, William (1564-1616)	<a href="#">Lear király /</a>	2011	Könyv	<a href="#">Részletek</a>
34.	<input type="checkbox"/>	Shakespeare, William (1564-1616)	<a href="#">Lear király /</a>	2010	Könyv	<a href="#">Részletek</a>
35.	<input type="checkbox"/>	Shakespeare, William (1564-1616)	<a href="#">Lear király /</a>	2008	Könyv része	<a href="#">Részletek</a>

5. ábra. Egy hagyományos OPAC „Lear király” keresőkérésre adott találati listájának részlete

Az OPAC, a számítógépesített katalógus ilyen értelemben tehát a cédulakatalógus e-könyv változatának felel meg. Itt tartunk most. Az igazi változást, a paradigmaváltást a cédulakatalógus hálózati kiadása jelentené, azaz szemantikus katalógus építése, mert nem járunk messze a valóságtól, ha azt mondjuk, hogy a szemantikus katalógus nem más, mint a könyvtári katalógus hálózati kiadása. Ha ugyanis sorra vesszük a hálózati könyvek jellegzetességeit, azt találjuk, hogy a szemantikus katalógus valóban hordozza a hálózati könyv valamennyi fontos tulajdonságát. Tudásgráfra épül, tehát strukturálisan szemcsés és hálózati, *crowdsourcing*- és *citizen science* technikákat alkalmaz, tehát közösségi és szerkeszthető/bővülő tartalommal rendelkezik. A szemantikus katalógust joggal nevezhetjük tehát hálózati könyvnek, a hagyományos könyvtári cédulakatalógusok hálózati kiadásának.



## **8. Konklúzió**

Szemantikus katalógus építése, azaz a hagyományos könyvtári katalógusok hálózati kiadásának elkészítése digitális bölcsészeti vállalkozás, mégpedig olyan, amelyben a könyvészetben jártas könyvtárosok közvetlen, aktív, alkotó részvétele nélkülözhetetlen. Amikor tehát a könyvtárak jövőbeli szerepéről töprengünk, egy digitális bölcsészeti műhely képe is felderenghet gondolataink között.

### **Semantic Catalogue as Digital Humanities**

In the dialogue on the relationship between libraries and the digital humanities, the former is commonly referred to as the creator of the conditions for the latter. The library provides and takes care of the digital ecosystem in which digital humanities projects can take place. Libraries can be directly and creatively involved in digital humanities undertaking since not only can philology, textology, linguistics, literature or history become digital but also the discipline of descriptive enumeration and systematic description of books: the bibliography. This paper argues that the networked edition of the traditional library card catalogue (building a semantic catalogue serving as a domain knowledge system) can be regarded as digital bibliography.

Keywords:

digital humanities, digital bibliography, semantic catalogue, knowledge graph, citizen science, networked book



## Ring Orsolya

*Társadalomtudományi Kutatóközpont, Politikatudományi Intézet*

*CSS-RECENS Kutatócsoport*

ring.orsolya@tk.hu

## Kiss László

*Társadalomtudományi Kutatóközpont*

*CSS-RECENS Kutatócsoport*

kiss.laszlo@tk.hu

# Agrárpolitikai kihívások és jogszabályalkotás a korai Kádár-korban Történeti források szövegelemzése és szövegbányászati vizsgálata\*

Tanulmányunkban a magyarországi kollektivizálás második hullámát egy, a társadalomtudományban használatos módszer adaptálásának segítségével vizsgáljuk, rámutatva, hogy a nagy adatbázisok elemzése milyen nagyobb, általánosabb történettudományi összefüggések levonására kínál lehetőségeket. Majd ugyanezt a kutatási kérdéskört egy szövegbányászati módszer, az úgynevezett topikmodellezés alkalmazásával elemezzük. Írásunk célja annak illusztrálása, hogy az egyre nagyobb számban elérhető digitális gyűjteményeknek a hazai történettudomány számára még részben újfajta módszerekkel történő elemzése milyen lehetőségeket és korlátokat rejt magában. Elemzésünk a korai Kádár-korszak agrárpolitikai jogszabályalkotásának vizsgálatát végzi el, arra a fő kutatási kérdésre keresve a választ, hogy milyen módon tükröződik vissza a jogszabályalkotásban a korszak gazdaságpolitikája, milyen módon reagált a jogszabályalkotás a korszak gazdaság- és társadalompolitikai változásaival kapcsolatban felmerülő szabályozási igényekre. A témaalapú osztályozás mellett szövegeinket szövegbányászati módszerekkel is vizsgálat alá vesszük, elemzésünk harmadik részében pedig kísérletet teszünk a két módszer kombinálására – azzal a céllal, hogy összefüggést találjunk a CAP-kódkönyvben meghatározott közpolitikai topikok, konkrétan az agrárpolitikai altopikok, valamint a topikmodellezés során kapott témák között.

Kulcsszavak:

digitális bölcsészet, történettudomány, szövegbányászat, osztályozás, topikmodellezés

\* A kutatást az FK-131826 témaszámú NKFIH projekt (Az 1945 és 1989 közötti sajtó társadalomtörténeti elemzése a természetesnyelv-feldolgozás [NLP] módszereivel) támogatta.



## 1. Bevezetés

Napjainkra hatalmas mennyiségű történeti forrásanyag vált elérhetővé és kereshetővé különböző digitális gyűjteményekben, így például az Arcanum Digitális Tudománytárban,<sup>1</sup> a Hungaricana<sup>2</sup> vagy az Adatbázisok Online<sup>3</sup> felületén. A digitalizált szövegek és az ezekhez rendelt metaadatok elemzése nagyban megkönnyíti a forrásokhoz való hozzáférést, de egyben új kihívásokat is támaszt, hiszen lehetőséget ad a *big data/big text* jellegű megközelítés történettudományi kutatásokra történő kiterjesztésére, amihez azonban a források újfajta megközelítése, feldolgozásukhoz pedig újfajta módszertan szükséges.

Gyáni Gábor szerint az állam és a politika múltját vizsgáló történetírás a 20. század második-harmadik évtizedéig tisztán narratív volt, alkalmazói egyáltalán nem éltek a kvantifikáció eszközével. A két világháború között következett be az, az *Annales* folyóirat vezető történészei által kezdeményezett paradigmaváltás, aminek következtében a történetírás szakított a szorosan vett filológiai, szövegkritikai módszertannal, és elkezdte kihasználni a mérésben rejlő elemzési lehetőségeket, elsődlegesen a gazdaság- és társadalomtörténet-írás területén a különböző statisztikai adatsorok elemzésbe való bevonásával. A kimondottan kvantitatív történeti kutatási programok először az 1970-es években öltöttek komolyabb méreteket, több nagy kutatócsoport is alakult a 19–20. századi, nagy tömegű, formalizálható adatokat magukban rejtő forrásanyagok gépi feldolgozására.<sup>4</sup>

Jean Marczewski a kvantitatív módszereknek három gyakorlatát írta le. Az első az a megszokott történetírói gyakorlat, amely a történelmi események leírását – illusztrációképpen – statisztikai adatokkal egészíti ki; a második az úgynevezett kliometrikus történetírás, ahol a történész közgazdasági magyarázó modelleket tesztl történeti anyagon.<sup>5</sup> Az ezt a módszert alkalmazó történész modelleket szerkeszt, melyekkel a történeti valóságot értelmezi, bemutatja a változók közötti lehetséges viszonyokat, megkeresi a változóknak a történeti múltban megfelelő értékeket, és az ebből következő függő változók értékét összeveti a változók tényleges értékével. A harmadik az a fajta kvantitatív történetírás, amely az időben visszamenőlegesen hozzáférhető nemzetgazdasági mérlegekre és statisztikákra épít, de nem csupán retrospektív mérlegeket konstruál, hanem „lényeges szerepet tulajdonít az adott időben és adott helyen lejátszódott történeti eseményeknek is”.<sup>6</sup>

<sup>1</sup> Hozzáférés: 2021.02.10, <https://adtpplus.arcanum.hu/hu/>.

<sup>2</sup> Hozzáférés: 2021.02.10, <https://hungaricana.hu/hu/>.

<sup>3</sup> Hozzáférés: 2021.02.10, <https://adatbazisokononline.hu/>.

<sup>4</sup> Gyáni Gábor, „Egy új történeti forrás: az adatbank,” *Történelmi Szemle* 24, 1. sz. (1981): 94–99, 94; Gyáni Gábor, „A statisztika mint konstruált és/vagy objektív történeti adat,” *Statisztikai Szemle* 96, 3. sz. (2018): 245–254, 246, <http://dx.doi.org/10.20311/stat2018.03.hu0245>.

<sup>5</sup> Eddie Scott M., *Ami „köztudott”, az igaz is? Bevezetés a kliometrikus történetírás gondolkodásmódjába* (Debrecen: Csokonai Kiadó, 1996).

<sup>6</sup> Jan Marczewski, „Kvantitatív történetírás,” in *Történetelméleti és módszertani tanulmányok*, szerk. Glatz Ferenc (Budapest: Gondolat Kiadó, 1977), 291–303, 294. Gyáni, „A statisztika,” 249–250.

A fenti igen vázlatos áttekintés is illusztrálja, hogy a kvantitatív történettudomány elsősorban a forrásokban rejlő strukturált, számszerű(síthető) adatok – statisztikák, összeírások stb. – elemzésében ragadható meg. Az 1970-es években a hazai történettudományban is megkezdődött a diskurzus a számítógépek kutatási alkalmazásainak lehetőségeiről. Már ekkor felismerték, hogy a számítógép használata nemcsak olyan források feldolgozását teszi lehetővé, amire manuálisan nem nyílta lehetőség, de egyben a forrásokból kinyerhető információk körét is bővíti.<sup>7</sup> A számítógépek és a történettudomány kapcsolatáról folyó diskurzus párhuzamosan zajlott és összekapcsolódott a történettudományi kutatások kvantifikálásáról szóló eszmecserevel.<sup>8</sup>

A társadalomtudományokban az 1980–1990-es évek fordulóján jelentek meg az első olyan statisztikai eljárások, amelyek nagyobb szövegtörzseket voltak képesek komplex módon elemezni, de a módszer használatának felfutása csak a 2000-es években indult meg.<sup>9</sup> Az automatizált szövegelemzési eljárások megjelentek más tudományterületeken, így például az irodalomtudományban<sup>10</sup> és a történettudományban is.<sup>11</sup> Bár a digitalizáció révén egyre több számítógéppel is elemezhető, strukturálatlan történeti adathalmaz (dokumentum) áll rendelkezésünkre, amely alkalmas mind a számítógéppel támogatott kvalitatív, mind a kvantitatív szövegelemzésre, a nagy szövegtörzsek számítógépes elemzésének elterjedését nagyban nehezítette és nehezíti, hogy az elemzendő szövegek nem digitálisan keletkeztek, azokat visszamenőleg digitalizálni kell, az így kapott szövegek minősége pedig gyakran csak sok előkészítő, tisztító munka után alkalmas az automatizált szövegelemzésre. Bár az OCR-szoftverek az utóbbi években jelentős fejlődésen mentek keresztül, az eredetileg sokszor rossz minőségű papíron keletkezett, vagy nem megfelelő felbontásban szkennelt

<sup>7</sup> Gyáni, „Egy új történeti forrás,” 94. Granasztói György, „A számítógépek a történettudományban,” *Történelmi Szemle* 15, 1–2. sz. (1972): 29–47, 36; lásd továbbá „Számítógépek és a történettudomány,” *Történelmi Szemle* 15, 3–4. sz. (1972): 526–583 – lásd Katus László *A számítógép alkalmazásáról a kapitalista korszak kutatásában* című előadásáról szóló részt: 535–536.

<sup>8</sup> Benda Gyula, „Kvantifikáció és történelem. (Hozzászólás Granasztói György *A történész és a mérés – egy modell korlátai* című írásához),” *Történelmi Szemle* 22, 1. sz. (1979): 105–108.

<sup>9</sup> Gary Miner, John Elder IV, Andrew Fast, Thomas Hill, Robert Nisbet and Dursun Delen, *Practical Text Mining and Statistical Analysis for Non-Structured Text Data Applications* (Cambridge: Academic Press, 2012), <https://doi.org/10.1016/C2010-0-66188-8>; Bing Liu, *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions* (Cambridge: Cambridge University Press, 2015), Jun Zhao, Kang Liu, and Liheng Xu, „*Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Bing Liu (University of Illinois at Chicago) Cambridge University Press, 2015, 381 pp.; hardcover, ISBN 9781107017894, \$80,” *Computational Linguistics* 42, 3. sz. (2016): 595–598, [https://doi.org/10.1162/COLI\\_r\\_00259](https://doi.org/10.1162/COLI_r_00259); Németh Renáta, Katona Eszter és Kmetty Zoltán, „Az automatizált szövegelemzés perspektívája a társadalomtudományokban,” *Szociológiai Szemle* 30, 1 (2020): 44–62, 45.

<sup>10</sup> Franco Moretti, *Distant Reading* (New York–London: Verso Books, 2013); Péter Róbert, „A Big Data kihívás és lehetőség a bölcsészettudományokban: Digitális szövegek és metaadatok távoli olvasása,” *Magyar Tudomány* 177, 11. sz. (2016): 1323–1330.

<sup>11</sup> Shawn Graham, Ian Milligran and Scott Weingart, *Exploring Big Historical Data: The Historian's Macroscopic* (London: Imperial College Press, 2015), <https://doi.org/10.1142/p981>; B. Alex, C. Grover, E. Klein, C. Llewellyn and R. Tobin, „User-Driven Text Mining of Historical Text,” in *Working with Text: Tools, Techniques and Approaches for Text Mining*, eds. Emma L. Tonkin and Gregory J. L. Tourte (Oxford: Chandos Publishing, 2016), 209–230, <https://doi.org/10.1016/B978-1-84334-749-1.00009-3>.

dokumentumok szövegbányászati elemzésre való előkészítése még mindig jelentősen több időráfordítást igényel, mint amire a *webscraping* segítségével gyűjthető szövegek esetében van szükség.<sup>12</sup>

A kvantitatív (pl. statisztika) és kvalitatív (pl. szöveges dokumentum) forrástípusok, a kvantitatív és kvalitatív történeti elemzés mellett különbséget tehetünk a szövegek kvantitatív és kvalitatív elemzési módszertana között is. Kvalitatív módszertan (kvalitatív szöveg-, illetve tartalomelemzés, számítógéppel támogatott szövegelemzés) alkalmazása esetén a szöveget jegyzetekkel és kódokkal látjuk el, amit végezhetünk gépi támogatással vagy anélkül, de a szövegen magán nem végzünk beavatkozást.<sup>13</sup> Kvantitatív szövegelemzés (szövegbányászat, szöveganalitika) esetén viszont az eredendően kvalitatív információforrást kvantitatív módon képezzük le.<sup>14</sup>

A szöveges történeti források kvantitatív szövegbányászati és statisztikai módszerek használatával történő elemzése révén lehetőséget kaphatunk a szövegek közötti kapcsolatok, a bennük rejlő mintázatok feltárására és elemzésére, olyan módon, ahogyan azt a szövegek egyenkénti olvasásával nem tudnánk megtenni, s ezáltal új összefüggések feltárására nyílik lehetőségünk. Írásunk célja annak illusztrálása, hogy az egyre nagyobb számban elérhető digitális gyűjteményeknek a hazai történettudomány számára még részben újfajta módszerekkel történő elemzése milyen lehetőségeket és korlátokat rejt magában.<sup>15</sup>

Esettanulmányunk a magyarországi kollektivizálás második hullámát (1957–1963) járja körül, ami egy viszonylag jól feldolgozott téma; célunk éppen annak bemutatása, hogyan lehet új elemzési eszközökkel, más tudományágak által alkalmazott modellek segítségével tovább mélyíteni a folyamatról korábban alkotott képet. Az elemzés

<sup>12</sup> Daniel Lopresti, „Optical Character Recognition Errors and their Effects on Natural Language Processing,” *International Journal on Document Analysis and Recognition (IJ DAR)* 12, 3. sz. (2009): 141–151, <https://doi.org/10.1007/s10032-009-0094-8>.

<sup>13</sup> Roel Popping, *Computer-Assisted Text Analysis* (London–California–New Delhi: Sage, 2000), <https://dx.doi.org/10.4135/9781849208741>; Sebők Miklós, Zorigt Burtejin és Zágoni Bella, „A szövegek társadalomtudományi elemzése: Egy elméleti áttekintés,” in *Kvantitatív szövegelemzés és szövegbányászat a politikatudományban*, szerk. Sebők Miklós (Budapest: L'Harmattan Kiadó, 2016), 15–22, 20.

<sup>14</sup> Justin Grimmer and Brandon Stewart M., „Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts,” *Political Analysis* 21, 3. sz. (2013): 267–297, <https://doi.org/10.1093/pan/mps028>; Sebők, Zorigt és Zágoni, „A szövegek társadalomtudományi elemzése,” 20–21.

<sup>15</sup> Hasonló történeti szövegkorpuszok NLP (Natural Language Processing) módszerekkel történő elemzésével foglalkoznak például az alábbi munkák is: Martina Katalin Szabó, Orsolya Ring, Balázs Nagy, László Kiss, Júlia Koltai, Gábor Berend, László Vidács, Attila Gulyás and Zoltán Kmetty, „Exploring the Dynamic Changes of Key Concepts of the Hungarian Socialist Era with Natural Language Processing Methods,” *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 2020. szept. 23., <https://doi.org/10.1080/01615440.2020.1823289>; Zoltán Kmetty, Veronika Vincze, Dorottya Demszky, Orsolya Ring, Balázs Nagy and Martina Katalin Szabó, „Pártélet: A Hungarian Corpus of Propaganda Texts from the Hungarian Socialist Era,” in *Proceedings of the 12th Language Resources and Evaluation Conference*, eds. Nicoletta Calzolari et al., (Marseilles: European Language Resources Association, 2020), 2381–2388; Ring Orsolya, Kmetty Zoltán, Szabó Martina Katalin, Kiss László, Nagy Balázs és Vincze Veronika, „Kulcsfogalmak jelentésváltozása a Kádárkorszak politikai diskurzusában,” in *XVI. Magyar Számítógépes Nyelvészeti Konferencia*, szerk. Berend Gábor, Gosztolya Gábor és Vincze Veronika (Szeged: Szegedi Tudományegyetem, 2020), 333–342.

során a kollektivizálás mint történelmi eseménysor politikai vetületére koncentrálni mutatjuk be az agrárszektor átalakításának megjelenését a Kádár-korszak jogszabályalkotásában, az MSZMP Politikai Bizottsága napirendjében, kitérve a kollektivizálással kapcsolatos egyéb társadalom- és gazdaságpolitikai témák megjelenésére, felfutására, ezek összefüggéseire is.

## 2. Források és adatbázisok

A *Magyar Közlönyben* megjelent jogszabályok adatbázisát az Optijus anyagainak felhasználásával állítottuk össze.<sup>16</sup> A jogalkotás vizsgálata során a *Magyar Közlönyben* publikált nyilvános anyagok mellett a Minisztertanács által kiadott titkos határozatok kódolt adatbázisát is bevontuk az elemzésbe. Az MSZMP Politikai Bizottsága napirendi jegyzéke esetében a Magyar Nemzeti Levéltár napirendi jegyzékeit és adatbázisait használtuk, ez alapján épült fel az MSZMP PB napirendjének adatbázisa.<sup>17</sup>

Adatbázisaink alap metaadat-struktúrája a jó összehasonlíthatóság érdekében minden esetben ugyanazt a rendszert követi. Minden egyes adatbázisrekord egyedi azonosítót tartalmaz, valamint az adott rekord dátumát, illetve ha rendelkezésre áll, a korabeli egyedi azonosítószámát, egy részletes leíró címet és az adott rekord közpolitikai kódszámát.

A *Magyar Közlönyben* megjelent jogszabályok szövegbányászati elemzéséhez szövegeinket PDF-ből TXT-formátumra konvertáltuk, a visszakereshetőség érdekében a fájlok elnevezésében szerepeltettük a kódolt adatbázis „id” oszlopában lévő egyedi azonosítószámot.

## 3. Módszertan

Tanulmányunkban először kvalitatív módszertannal nagy adatbázisokon végzett kutatást ismertetünk, egy a társadalomtudományban használatos módszer adaptálására, és annak segítségével történettudományi kérdések megválaszolására teszünk kísérletet. Célunk annak bemutatása, hogy a nagy adatbázisok elemzése milyen nagyobb, általánosabb összefüggések levonására ad lehetőséget, hogyan használható arra, hogy megtaláljuk azokat a területeket, ahol érdemes a kutatás során mélyfúrásokat végezni, továbbá hogyan segítheti a történeti kutatásokat az eredmények számszerűsítésével. A választott módszertan egy deduktív eljárás, az osztályozás vagy klasszifikálás, amely során az elemezni kívánt szövegtörzsek elemeit előzetesen kialakított kategóriákba

<sup>16</sup> Hozzáférés: 2021.02.10, <https://optijus.hu/jogi-informacios-szolgalattasok/optijus>.

<sup>17</sup> A mostani elemzésbe bevont jogszabályok a teljes korpusznak csak egy részhalmazát teszik ki. A jelenleg rendelkezésre álló teljes adatbázis a *Magyar Közlönyben* megjelent 1957 és 1989 közötti jogszabályok esetében 13918, a Minisztertanács titkos határozatainak esetében 5228 kódolt rekordot tartalmaz. A titkos határozatok sora nem teljes, mivel ezek csak néhány éve veszítették el titkos minőségüket, így a belőlük való adatbázis-építés jelenleg is folyamatban van, jelenleg az 1957 és 1967 közötti titkos jogszabályok feldolgozása történt meg. Az 1957 és 1989 közötti időszak MSZMP PB adatbázisa 19843 kódolt napirendi pontot tartalmaz.

rendeztük, azaz előre tisztában voltunk a sokaság rendezésének alapjául szolgáló jellemzőkkel.<sup>18</sup>

Elemzésünk során a politikatudományban használt, úgynevezett CAP-kódolás (Comparative Agendas Project) módszerét alkalmaztuk, hogy meghatározzuk adatbázisainkban az egyes közpolitikai témákat, és számszerűsítsük előfordulásukat. Frank R. Baumgartner és Bryan D. Jones amerikai politológusok a közpolitikai változás mintázatait vizsgálva figyeltek fel arra, hogy ezek a változások általában fokozatosan következnek be, de időnként felgyorsulnak és a kormányzat látszólag váratlanul kezd reformokba egy-egy közpolitikai területen. Baumgartner és Jones szerint a közpolitikai változás részben magából a politikai-közpolitikai rendszer működéséből ered mivel a közpolitikai aktorok (pártok, civil szervezetek) időnként képesek arra, hogy egy-egy kérdést a közpolitikai napirendekre felvigyenek, és ha eléri, hogy a nagypolitika figyelme is ráirányuljon az adott problémára, akkor ezáltal akár jelentős közpolitikai változásokat idézhetnek elő.<sup>19</sup>

Baumgartner és Jones elemzésükhöz kialakítottak egy kódkönyvet, amelyben eredetileg 19 közpolitikai főcsoportkódot (pl. makrogazdasági politika) és ezeken belül összesen 225 altémát (pl. költségvetési politika, monetáris politika stb.) határoztak meg. A kódok kizárólagosak, tehát minden megfigyelés csak egy kódot kaphat. Ez a kódolás lehetővé teszi például annak vizsgálatát, hogy egy adott közpolitikai témáról mikor beszélt sokat a sajtó, mikor tartották az emberek is fontosnak, és mikor született belőle jogszabály.<sup>20</sup> Baumgartner és Jones projektje (Comparative Agendas Project)<sup>21</sup> mára a nemzetközi politikatudomány egyik jelentős adatbázisprojektjévé nőtte ki magát, amelyben az első kelet-közép-európai résztvevő a Hungarian Comparative Agendas Project<sup>22</sup> lett. Ennek célja a magyar közpolitika dinamikájának feltárása, amihez konzisztens adatbázisokat hoznak létre a magyar közpolitikai folyamatok meghatározó aktorai – így a parlament, a média és a közvélemény szereplői – számára meghatározó ügyekről. A módszer megbízhatóságát az egységes és pontosan lefedett kódolási szabályok garantálják, amelyek révén a különböző közpolitikai területek prioritásainak változásai időben és területenként is nyomon követhetők.<sup>23</sup>

Az osztályozást a nemzetközi CAP-projekt módszertana alapján kettős vak kézi kódolással végeztük, azaz minden esetben két kutató egymástól függetlenül határozta meg, hogy az adott megfigyelés melyik közpolitikai topikba vagy altopikba tartozik.

<sup>18</sup> Szövegbányászat, szerk. Tikk Domonkos (Budapest: Typotext Kiadó, 2007); Molnár Csaba, „Osztályozás (klasszifikáció),” in Sebők, *Kvantitatív szövegelemzés*, 65–72.

<sup>19</sup> Boda Zsolt, „A közpolitikai változás elméletei és a megszakított egyensúly,” in *A magyar közpolitikai napirend: Elméleti alapok, empirikus eredmények*, szerk. Boda Zsolt és Sebők Miklós (Budapest: MTA TK PTI, 2018), 15–23.

<sup>20</sup> Frank R. Baumgartner and Bryan D. Jones, *Agendas and Instability in American Politics* (Chicago: University of Chicago Press, 2010); Boda Zsolt, „A közpolitikai változás elméletei és a megszakított egyensúly,” in Boda és Sebők, *A magyar közpolitikai napirend*, 15–23; Sebők Miklós, „A magyar Comparative Agendas Project módszertana,” in *uo.*, 23–30.

<sup>21</sup> Hozzáférés: 2021.02.10, <https://www.comparativeagendas.net/>.

<sup>22</sup> Hozzáférés: 2021.02.10, <https://cap.tk.mta.hu/>.

<sup>23</sup> Boda Zsolt és Sebők Miklós, „Előszó: a Hungarian Comparative Agendas Project bemutatása,” *Politikatudományi Szemle* 24, 4. sz. (2015): 33–40.



Miután a két független kódolást összevetettük, a vitás esetekben konszenzusos kódokat osztottunk ki.<sup>24</sup>

Ezt követően ugyanazt a kutatási kérdéskört kvantitatív módszertannal is megvizsgáltuk az egyik legelterjedtebb szövegbányászati módszer, az úgynevezett csoportosítás vagy klaszterezés és azon belül a topikmodellezés alkalmazásával. Ebben az esetben a korpusz elemeinek rendezéséhez nem határozzunk meg előzetesen csoportokat. A topikmodellezés olyan felügyelet nélküli tanulási módszer, amelynek során az alkalmazott algoritmus a dokumentum tulajdonságait és a modell becsléseit felhasználva hoz létre különböző kategóriákat, amelyekhez később hozzárendeli a szöveget.<sup>25</sup> Az általunk alkalmazott topikmodellezési eljárás, a látens Dirichlet-allokáció (LDA) alapja az a feltételezés, hogy minden korpusz topikok/témák keverékéből áll, ezen témák pedig statisztikailag a korpusz szókészlete valószínűségi függvényeinek (eloszlásának) tekinthetők.<sup>26</sup> Az LDA a korpusz dokumentumainak csoportosítása során az egyes dokumentumokhoz topikszavakat rendel, a topikok megbecsléséhez pedig a szavak együttes megjelenését vizsgálja a dokumentum egészében. Az LDA algoritmusnak előzetesen meg kell adni a keresett klaszterek (azaz a keresett topikok) számát, ezt követően a dokumentumhalmazban szereplő szavak eloszlása alapján az algoritmus azonosítja a kulcsszavakat, amelyek eloszlása kirajzolja a topikokat.<sup>27</sup>

A klaszterezés vagy csoportosítás egy induktív kategorizálás, ami akkor hasznos, amikor nem állnak a kutató rendelkezésére előzetesen ismert csoportok, amelyek szerint a vizsgált dokumentumokat rendezni tudná. Hiszen ebben az esetben a korpusz elemeinek rendezéséhez nem határozzunk meg előzetesen csoportokat, hanem az eljárás során olyan különálló csoportokat hozunk létre a dokumentumokból, amelyek tagjai valamilyen szempontból hasonlítanak egymásra. A csoportosítás legfőbb célja, hogy az egy csoportba kerülő szövegek minél inkább hasonlítsanak egymásra, miközben a különböző csoportba kerülők minél inkább eltérjenek egymástól. Azaz klaszterezésnél nem egy-egy szöveg jellemzőire vagyunk kíváncsiak, hanem arra, hogy a szövegek egy-egy csoportja milyen hasonlóságokkal bír.<sup>28</sup>

Az általunk választott osztályozás és klaszterezés természetesen egymástól függetlenül is alkalmazható eljárások, de jól ki is egészíthetik egymást.

<sup>24</sup> A CAP-kódolás nemcsak kézzel, hanem felügyelt gépi tanulással is végezhető, 1990 utáni magyar nyelvű dokumentumokon tanítva jelenleg is használatban van egy felhőalapú klasszifikációs eljárás, amely nagy hatékonysággal képes a CAP (fő)kategóriáinak osztályozására. Miklós Sebők and Zoltán Kacsuk, „The Multiclass Classification of Newspaper Articles with Machine Learning: The Hybrid Binary Snowball Approach,” *Political Analysis*, 2020. nov. 11., 1–14, <https://doi.org/10.1017/pa.2020.27>.

<sup>25</sup> Grimmer and Stewart, „Text as Data,” 15.

<sup>26</sup> Németh, Katona és Kmetty, „Az automatizált szövegelemzés,” 55; David M. Blei, Andrew Y. Ng and Michael I. Jordan, „Latent Dirichlet Allocation,” *Journal of Machine Learning Research* 3, 1. sz. (2003): 993–1022.

<sup>27</sup> Zorigt Burtejin, „Csoportosítás (klaszterezés),” in Sebők, *Kvantitatív szövegelemzés*, 85–101, 98.

<sup>28</sup> Tikk, *Szövegbányászat*, 145.

## 4. Elemzés

### 4.1. Az agrárszektor helyzete a korai Kádár-korszakban

Az 1956-os forradalom eredményeként 1957. január 1-ig az 1956 októberében működő tsz-ek 62%-a felbomlott.<sup>29</sup> A forradalom leverését követően hatalomra kerülő Kádár János érzékelve az agrárirányítás Rákosi-korszakbeli csődjét, új irányvonalat határozott meg. A mezőgazdasági termelőket sújtó begyűjtési rendszer 1956. novemberi „hivatalos” eltörlése ugyan még inkább csak jelképes politikai aktus volt (a begyűjtési rendszer eddigre már rég összeomlott), az MSZMP 1957 nyarán közzétett Agrárpolitikai Tézisei azonban már részletesen bemutatták a meghirdetésre kerülő új mezőgazdasági és szövetkezeti politikát.

1957-ben a mezőgazdasági termények zömét még magángazdaságok állították elő. Az 1959-ben megindult kollektivizálás eredményeként 1961-re az ország mezőgazdaságilag hasznosítható területének 80%-a került a tsz-ek használatába.<sup>30</sup> Az egyéni gazdaságok száma 165000-re csökkent, s a helyüket 271 állami, és körülbelül 4200 termelőszövetkezeti gazdaság vette át.<sup>31</sup> A számok mögött látható a hagyományos paraszti társadalom erőszakos (és csaknem teljes) felszámolása, a parasztság termelőszövetkezetekbe kényszerítése.<sup>32</sup>

A jogalkotás új feladata lett az állami tulajdonban lévő állami gazdaságok működésének szabályozása és a termelőszövetkezeti működés irányítása – ennek fő eszközei a tervezés- és támogatáspolitikai, a beruházások irányítása, az ár- és hitelpolitika, valamint a felvásárlási rendszer kiépítése. Az első átfogó szabályozást az 1959. évi 7. törvényerejű rendelet a mezőgazdasági termelőszövetkezetekről és a termelőszövetkezeti csoportokról (márc. 19.) adta.

A legfontosabb feladat a mezőgazdaság beszolgáltatási kötelezettségének visszafogásából adódóan az agrártermékek állami felvásárlásának, kvázi-kereskedelmének szabályozása lett. Az állam, illetve a továbbra is államilag szervezett, központosított élelmiszer-gazdaság csak úgy tudott agrártermékhez, alapanyaghoz jutni, ha valamilyen árrendszer bevezetésével és alkalmazásával megvásárolta azt a termelőktől – az egyénileg gazdálkodóktól, illetve a szövetkezetektől.<sup>33</sup> Közvetlen állami irányítás alatt, centralizálva működtek a gépállomások is, amelyek szerepe a szövetkezetek saját gépekkel való ellátásának lehetővé tételével párhuzamosan fokozatosan csökkent – ezen területek működésének újraszabályozása ugyancsak a korszak jogalkotásának fő feladata lett.<sup>34</sup>

<sup>29</sup> Valuch Tibor, „Agrárkérdések és a magyar falu 1956–57-ben,” in *Évkönyv VIII.*, szerk. Kőrösi Zsuzsanna, Ständeisky Éva és Rainer M. János (Budapest: 1956-os Intézet, 2000), 286–302, 300, 303.

<sup>30</sup> Fukker Valéria, „A pártoló tag – epizód a tsz-szervezések történetéből,” *ArchivNet* 6, 2. sz. (2006), [https://archivnet.hu/politika/a\\_partolo\\_tag\\_epizod\\_a\\_tszszervezesek\\_tortenetebol.html](https://archivnet.hu/politika/a_partolo_tag_epizod_a_tszszervezesek_tortenetebol.html).

<sup>31</sup> Varga Zsuzsanna, „Az agrárpia rehabilitálása Magyarországon 1956 után,” *Metszetek* 2–3. sz. (2012): 83–92, 88.

<sup>32</sup> Vö. Ö. Kovács József, *A paraszti társadalom felszámolása a kommunista diktatúrában: A vidéki Magyarország politikai társadalomtörténete 1945–1965* (Budapest: Korall Kiadó, 2012).

<sup>33</sup> Varga, „Az agrárpia rehabilitálása,” 87.

<sup>34</sup> Juhász Pál, „Agrárpia, kisüzem, nagyüzem: (Agrárszociológiai tanulmányok),” *Medvetánc* 2, 1. sz. (1982): 117–139, 120.

Az élelmiszer-ellátás biztosítása, a lakossági fogyasztás biztosítása, és lehetséges volumenének növelése komplex politikai kérdésként jelentkezett – a megkezdődő kádári konszolidációs folyamat egyik alapelemévé vált. A kádári életszínvonal-politika fontos eleme volt az első szándékból a termelőszövetkezeti parasztság fogyasztási színvonalának emelését célzó háztáji gazdálkodás bevezetése. A háztáji már rövidtávon komoly hatást ért el a termelőszövetkezeti parasztság életszínvonalának növelésében, kontraproduktív volt ugyanakkor maguknak a termelőszövetkezeteknek a működése szempontjából. A termelőszövetkezetek nem megfelelő jövedelemelosztása, a szakemberhiányból fakadó sok vezetési probléma, de vélhetően az erőszakos kollektivizálás „utóhatása”, a tsz-szel mint termelési egységgel való ellenségesség is hozzájárult ahhoz az érdektelenséghez, ami a tsz-ekben végzett munkát jellemezte. A szövetkezetek működése akadozott, az élelmiszer-ellátásban betöltött szerepük elégtelen volt. Már 1959-től kezdve megjelent a „gyengén teljesítő” szövetkezetek megerősítésének, a szövetkezeti működés megszilárdításának jogszabályi igénye, amely összekapcsolódott a termelőszövetkezetek szakemberellátásának fokozásával.

1961 végén az MSZMP Politikai Bizottsága átfogó agrárpolitikai reformmunkába kezdett, amely három nagy területre terjedt ki: az agrárszektor új ár-, adó- és pénzügyi rendszerének kidolgozására, a mezőgazdasági irányítás felülvizsgálatára és az új termelőszövetkezeti törvény megalkotására. Bár 1963-ra el is készültek a mezőgazdasági reformjavaslatok, a bevezetésüket a gazdaságban jelentkező problémák miatt 1963–64 fordulóján elhalasztották, és csak 1967-ben született meg az átfogó újraszabályozást tartalmazó – 1967. évi III. törvény a mezőgazdasági termelőszövetkezetekről, majd az 1967. évi IV. törvény a földtulajdon és földhasználat továbbfejlesztéséről.<sup>35</sup>

#### 4.2. Osztályozás

A szocialista korszak(ok) igazgatását a megkettőzött struktúra, a párt- és az államigazgatási irányító szervek egymás melletti, egymás tevékenységét átfedő, alapvetően a pártszervek dominanciája mellett zajló tevékenység jellemezte. A Rákosi- és a Kádár-korszak igazgatási szabályozásának, ezen belül jogforrásainak rendszere meglehetősen bonyolult – a jogszabályok körét a törvények, törvényerejű rendeletek mellett országgyűlési, miniszteri, minisztertanácsi rendeletek és határozatok, utasítások sokasága alkotta.

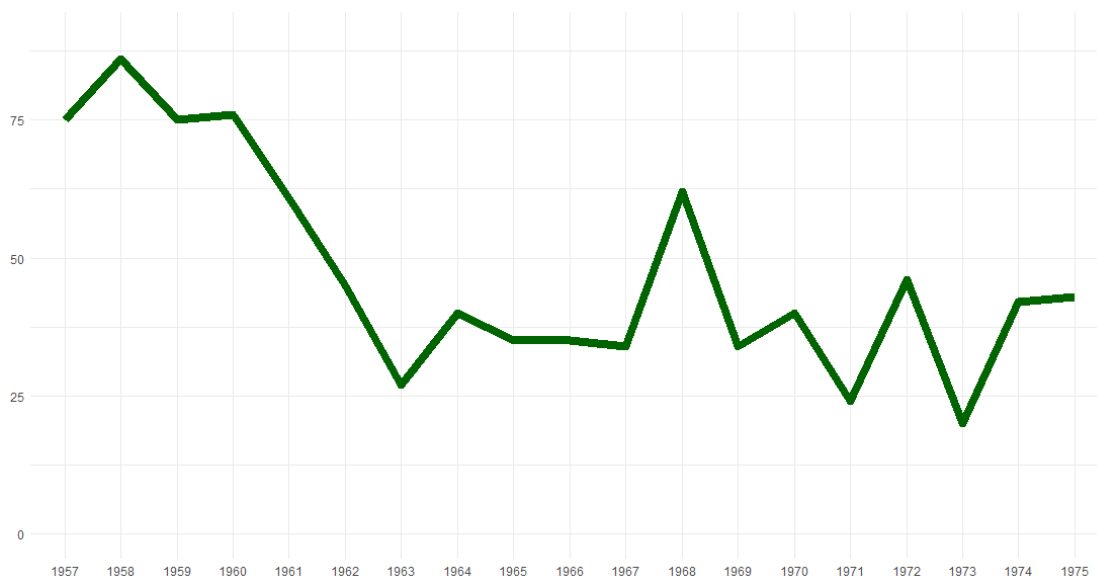
A jogforrások hierarchikus rendszerét a jogszabályok közzétételéről és hatálybalépéséről szóló 1954. évi 26. számú törvényrendelet alapozta meg, később pedig az 1974. évi 24. sz. törvényerejű rendelet szabályozta átfogó jelleggel. A szabályozás egyik központi kérdése volt a jogszabály kihirdetési kötelezettsége, vagy annak hiánya – a *Magyar Közlöny*ben kellett kihirdetni azokat a jogszabályokat, amelyek az állampolgárokra közvetlenül kötelezettséget róttak, az állampolgárok jogait vagy kötelességeit érintették, illetve amelyek állami szervekre, szövetkezetekre és egyéb gazdálkodó szervekre vonatkozóan általános érvényűek voltak.<sup>36</sup> A *Magyar Közlöny*ben kihirdetésre nem kötelezett, de titkosnak sem minősített jogszabályok egy részét az ugyancsak hi-

<sup>35</sup> Varga, „Az agrárpiac rehabilitálása,” 90.

<sup>36</sup> Ádám Antal, „Jogalkotás, jogforrások és a jogszabályok kihirdetése,” *Jogtudományi Közlöny* 30, 6. sz. (1975): 301–312, 303.

vatalos lapnak számító *Határozatok Tárában*, illetve a különböző ágazati értesítőkből, közlönyökben hirdették ki, sok jogszabálytípus esetében azonban felderíthetetlen volt a kihirdetési rend, sőt egyes, elvben nem titkosnak minősített jogszabályok semmilyen hivatalos fórumon nem kerültek nyilvánosságra.<sup>37</sup>

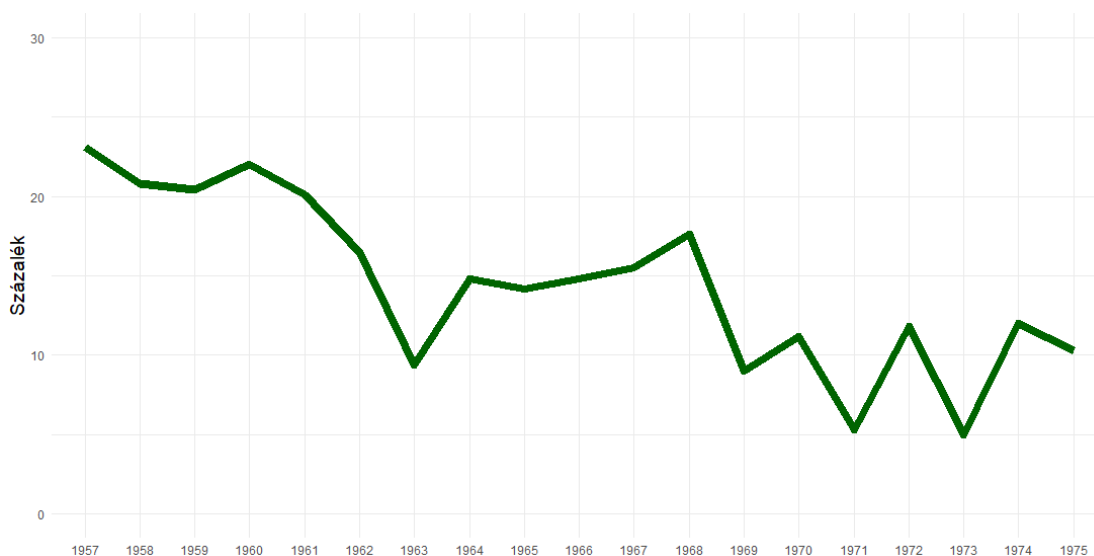
Ha a *Magyar Közlönyben* megjelent jogszabályok CAP-főtopikok szerinti megoszlását vizsgáljuk, jól láthatjuk, hogy a mezőgazdasági szektort érintő jogszabályok száma és aránya 1957 és 1961 között kiemelkedően magas. Az abszolút számokat tekintve csak az 1968-as „reformév” hozott hasonló értéket, az agrárszektor jogszabályainak az összes megjelent jogszabályhoz viszonyított aránya azonban még az „új gazdasági mechanizmus” bevezetésekor sem ért el olyan magas értéket, mint az 1950-es és 1960-as évek fordulóján.<sup>38</sup>



1. ábra. A *Magyar Közlönyben* megjelent agrárpolitikai jogszabályok száma, 1957–1975

<sup>37</sup> Ádám, „Jogalkotás, jogforrások,” 305.

<sup>38</sup> Bár elemzésünk az 1957 és 1963 közötti időszak jogszabályalkotását, valamint közpolitikai napirendjét vizsgálja, a kontextusba helyezés végett az agrárpolitikai jogszabályok, titkos határozatok és MSZMP PB napirendi pontok arányát az adott év teljes jogszabályi, határozati, illetve napirendi anyagán belül hosszabb időkeretben, 1957 és 1975 között mutatjuk be.



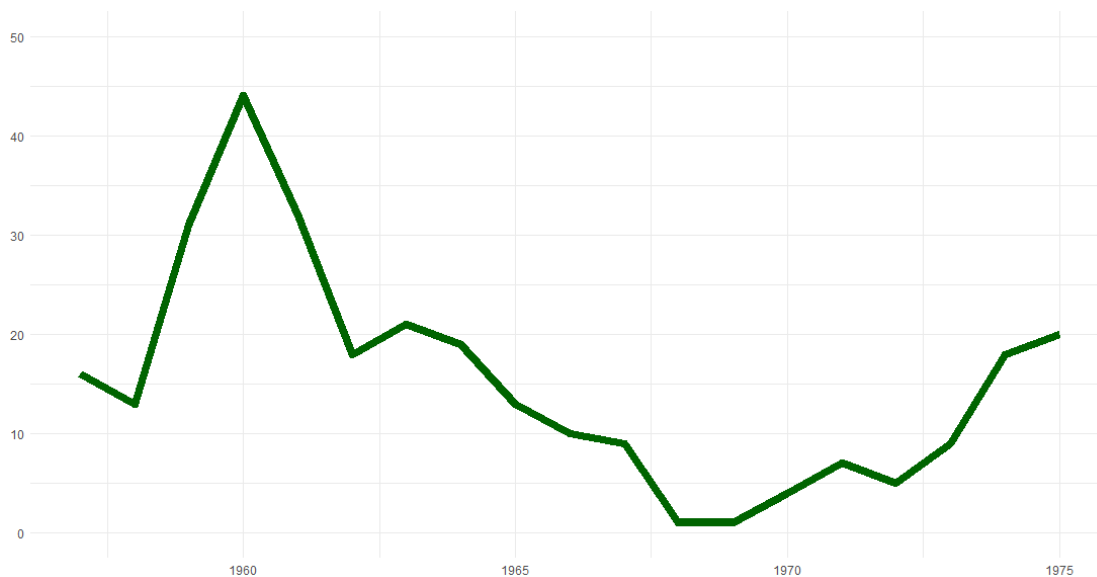
2. ábra. A Magyar Közlönyben megjelent agrárpolitikai jogszabályok összes megjelent jogszabályon belüli aránya (N=6494)

A mezőgazdaság kérdéskörének felértékelődését jól jelzi az MSZMP Politikai Bizottságának fokozott érdeklődése is. Az MSZMP által kiadott párthatározatok, pártutasítások, irányelvek stb. nem számítottak hivatalosan önálló jogforrásnak, a rendszer működéséből fakadóan azonban komoly jelentőséggel bírtak a tényleges jogszabályalkotás folyamatára. A párthatározatok a „demokratikus centralizmus” értelmében kötelező érvénnyel bírtak a pártszervekre, valamint a pártfegyelem révén a párt teljes tagságára nézve. „Miután a többség elfogadta a párthatározatot, annak pontos végrehajtása a pártélet feltétlen törvénye, amely kötelező minden párttagra nézve, bárhol dolgozik, és bármilyen tisztséget tölt be.”<sup>39</sup> A párthatározatok és az államigazgatási szervek által kiadott jogszabályok témák szerinti egymásra épülése ugyan nem feltétlenül valósult meg minden esetben, de kifejezetten gyakorinak számított, hogy a párthatározatokban, pártutasításokban megfogalmazott elvek, előírások egy későbbi hivatalos állami jogszabály szövegében is visszaköszöntek. Sajátos szereppel bírtak a Minisztertanács által alkotott, de kifejezetten titkosnak minősített határozatok. Azon túl, hogy egyes speciális témák szabályozását bizonyos időszakokban kizárólag titkos határozatok útján végezték, sok esetben itt is látható egyfajta egymásra épülés a nyilvános határozatokkal – a titkos határozatokban megalapozott, elindított folyamat végterméke sok esetben nyilvános jogszabály lett.

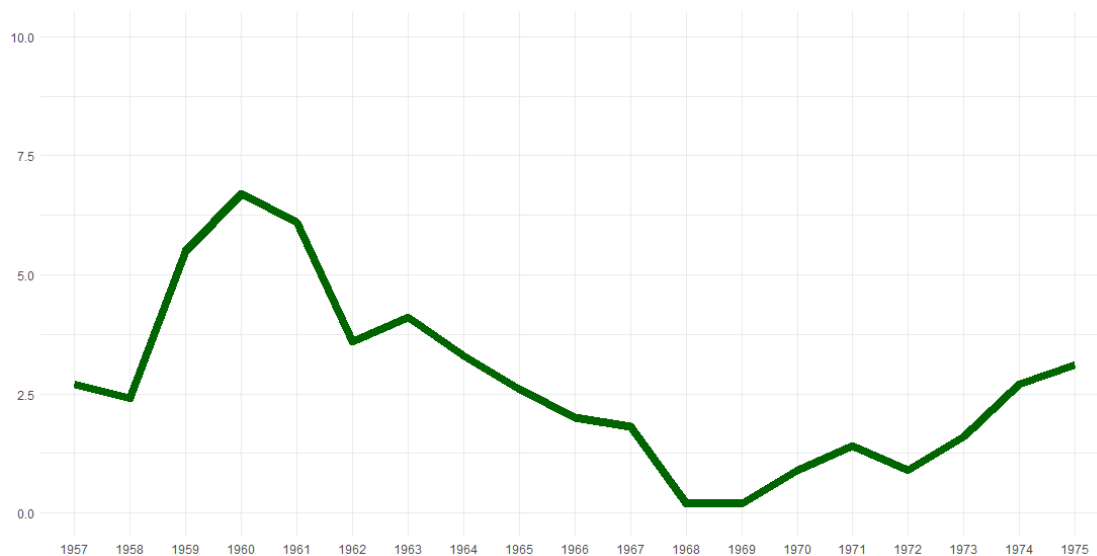
Az MSZMP által hozott párthatározatok száma 1957 és 1963 között 54 volt – ebből kifejezetten agrárpolitikával 15 határozat (27,8%) foglalkozott. Bár ez az arány kifejezetten magas, az abszolút szám relatíve alacsony, így a korpusz további bontásokra, altémakörökbe sorolásra nem alkalmas. A párthatározatok helyett így egy nagyobb korpuszon, az MSZMP Politikai Bizottság üléseinek napirendi pontjain végeztük el az elemzést. 1959 és 1961 között az ülések napirendi pontjai között évente 30–40 agrártéma szerepelt, míg más években ez az érték a 20-at sem nagyon érte el. A Politikai Bizottság tevékenységét ugyan a hatvanas évek elején is a kül- és a belpolitika

<sup>39</sup> Milko Balev, „A párthatározat,” *Béke és szocializmus* 25, 4. sz. (1982): 41–42.

dominálta, de a kollektivizálás időszakában 5–7%-ot ért el az összes napirendi ponton belül az agrárpolitika aránya.



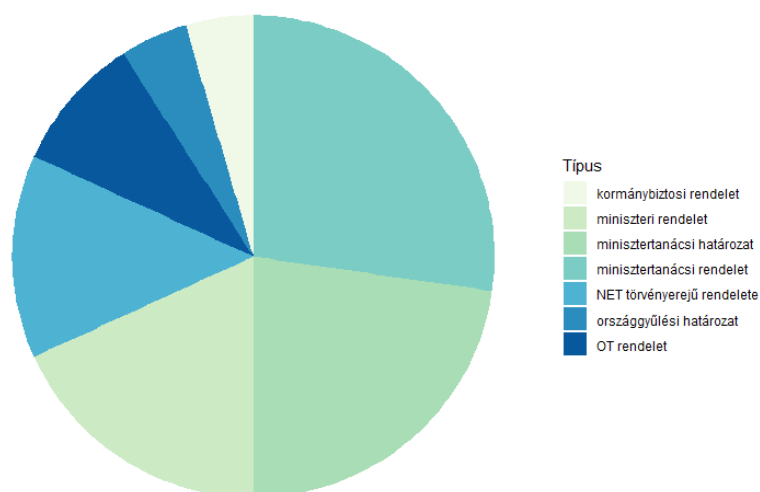
3. ábra. Az MSZMP Politikai Bizottság agrárpolitikai témájú napirendi pontjainak száma



4. ábra. Agrárpolitikai témájú napirendi pontok aránya az MSZMP Politikai Bizottságának napirendjén (N=10191)

A *Magyar Közlöny*ben megjelentetett, agrárpolitikai témájú jogszabályok jogforrások szerinti megoszlását vizsgálva a miniszteri rendeletek túlsúlya fedezhető fel. Az összes agrárpolitikai jogszabály mintegy háromnegyede az ágazatot felügyelő miniszter saját vagy más miniszterekkel közösen kiadott rendelete. A jogforrás-hierarchia magasabb

szintjén elhelyezkedő, az Elnöki Tanács által kiadott törvényerejű rendeletek a vonatkozó jogszabályok 7%-át, a Minisztertanács által alkotott rendeletek és határozatok pedig azok 9, illetve 8%-át tették ki. Az 1957 és 1963 közötti teljes *Magyar Közlönyben* megjelent joganyaggal összevetve az agrárpolitikai szabályozást, a területen a miniszteri rendeletek erőteljes felül-, míg a többi jelentősebb jogforrástípus közepes alulreprezentáltságot állapíthatjuk meg.



5. ábra. A *Magyar Közlönyben* megjelent agrárpolitikai témájú jogszabályok a jogforrás típusa szerint (N=445)

Az adatok alapján több vizsgálatra érdemes kérdést is megfogalmazhatunk: egyrészt milyen, a kollektivizáláshoz kapcsolódó folyamatok indokolták ezt a kiemelkedő méretű mezőgazdasági témájú jogszabályalkotást, másrészt milyen típusú jogszabályok születtek, az agrárszektor működésének mely szegmensei, milyen részterületei kerültek szabályozásra. A két kérdés összekapcsolásaként pedig megvizsgálhatjuk, milyen módon reagált a jogszabályalkotás a kollektivizálással kapcsolatban felmerülő szabályozási igényekre. A jogi szabályozási feladatok áttekintése után nézzük meg, hogyan alakult részleteiben az 1957 és 1963 között az agrárszektorra irányuló politikai figyelem, és milyen részterületekre fókuszáltak a szektort érintő jogszabályok. Az elemzéshez a CAP közpolitikai kódjait használjuk, ezek ugyanis a közpolitikai főtemakörök meghatározásán túl alkalmasak az altémakörök elkülönítésére is. Az „agrárpolitika” főtemakörön belül a CAP-kódkönyv kilenc altémakört tartalmaz:

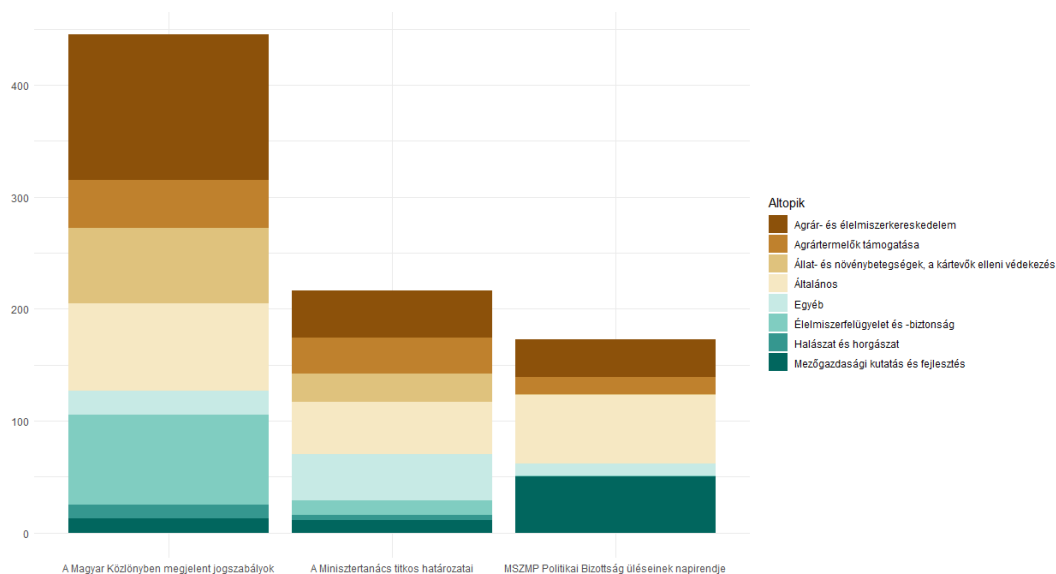
- általános (több altéma kombinációját tartalmazza)
- agrár- és élelmiszer-kereskedelem
- agrártermelők támogatása, intervenciós alapok és biztosítás
- élelmiszer-felügyelet és -biztonság
- agrármarketing és piackutatás
- állat- és növénybetegségek, a kártevők elleni védekezés
- halászat és horgászat

- mezőgazdasági kutatás és fejlesztés
- egyéb

Bár a CAP-kódrendszer nem minden esetben illeszthető könnyen a történeti kutatások szempontrendszeréhez, és az eltérő történelmi korszakok közpolitikájának tagolása, az egyes kategorizálendő elemek témakörökbe sorolása (különösen az altémakörök szintjén) is okozhat nehézségeket, mégis alapvetően alkalmas arra, hogy a közpolitikai figyelem alakulásának történeti elemzéséhez felhasználjuk.

Ha a teljes vizsgált időszakban nézzük az agrárpolitikai jogalkotást, valamint a Politikai Bizottság agrárpolitikával kapcsolatos napirendi pontjait, jól látható a különbség a jogszabályok és a napirendi pontok altémakörök szerinti megoszlása között. A párt és az állam agrárpolitikai tevékenysége eltérő irányultságú, más és más irányítási szegmensekre, területekre helyezi a hangsúlyt. A Politikai Bizottság napirendi pontjai között kiemelkedően sok foglalkozik általános agrárpolitikai témával (35%), valamint mezőgazdasági kutatással és fejlesztéssel (29%), tehát a testület munkáját inkább egyfajta makroszintű szemlélet jellemzi, míg a megjelent jogszabályok esetében az „operatív témák” dominálnak – köztük a párt érdeklődési köréből szinte teljességgel hiányzó állat- és növényvédelem, valamint az élelmiszer-felügyelet. A jogszabályi szint viszont alig érinti a láthatóan a párt hatáskörébe tartozó agrárfejlesztéseket.

Ha visszautalunk az előzőekben kifejtett politikai figyelmet indukáló, illetve jogszabályalkotást szükségessé tevő gazdasági és társadalmi folyamatokra, láthatjuk, hogy az agrártermékek kereskedelmével kapcsolatos árképzési, árusítási, felvásárlást szabályozó kérdések a jogszabályok, a minisztertanácsi titkos határozatok, valamint a Politikai Bizottság napirendi pontjai között is kiemelkedő arányban képviseltetik magukat. A nyilvános agrárpolitikai jogszabályok 30, a titkos határozatok 20%-a foglalkozik az agrárkereskedelemmel (ezen belül elsősorban az árszabályozással), és a Politikai Bizottság agrár tematikájú napirendi pontjainak is egyötödét teszik ki az ebbe a CAP-altémakörbe sorolható elemek.



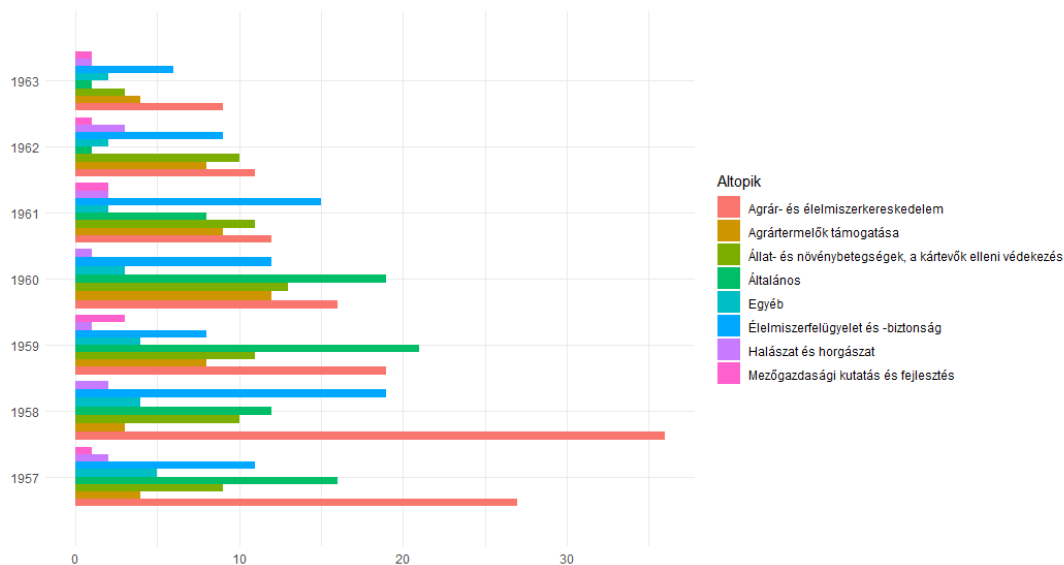
6. ábra. Agrárpolitikai témák altópikok szerinti megoszlása



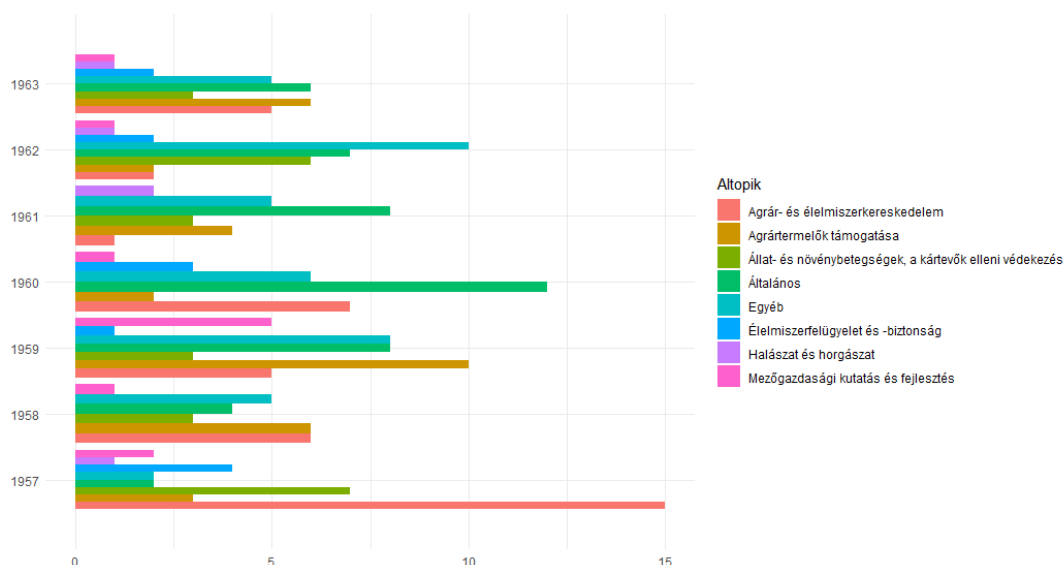
A teljes vizsgált korszakra vonatkozó összesített megoszlásoknál jobban mutatja a közpolitikai figyelem változását az egyes altémakörökbe sorolható elemek évenkénti előfordulása az agrárpolitikai jogszabályokon és napirendi pontokon belül.

A nyilvános jogszabályok, valamint a titkos minisztertanácsi határozatok adatsorai-ból kiválóan látszik, hogy az agrárkereskedelem szabályozása a korszak első éveiben adta a legtöbb munkát az államigazgatásnak – összhangban a korábbi történeti elemzés során említett ténnyel, amely szerint a begyűjtési rendszer kiiktatása következtében az igazgatás elsődleges feladatává vált az új árrendszer és az erre alapuló felvásárlási rendszer kialakítása. Az altémakör az egész vizsgált korszakban az agrárpolitika fontos területe maradt, az ide sorolható jogszabályok száma azonban az 1957-es és 1958-as csúcsevek után fokozatosan lecsökken.

A termelőszövetkezetek (és a termelők) megerősítésének politikai igénye is jól látszik az „agrártermelők támogatása” altémakörbe tartozó jogszabályok számának 1958–59-es megemelkedésében – a terület hangsúlyossá válása különösen a titkos minisztertanácsi határozatoknál szembeötlő, de 1959-ben a területhez kapcsolódó nyilvános határozatok száma is megnőtt. Ennél az altémakörnél a titkos határozatok és a kihirdetett jogszabályok egymásra épülésének sajátossága is megfigyelhető: 1958 elején a Minisztertanács a 3004/1958. A termelőszövetkezetek gazdasági megerősítését és fejlesztését szolgáló intézkedésekről titkos határozatban rendelte el a tsz-ek megerősítését szolgáló komplex (növénytermesztési és állattenyésztési fejlesztéseket, az értékesítés javítását, hitelezései kedvezményt, az állami szolgáltatások díjainak kedvezményét, szakemberek delegálását stb. tartalmazó) intézkedéssorozatot, egyben felhatalmazta, illetve megbízta az egyes szakterületi minisztereket a szükséges rendeletek, határozatok meghozatalával. A *Magyar Közlönyben* 1958 októberében jelent meg a titkos kormányhatározat kihirdetett végrehajtási utasítása, majd a következő év elejétől kerültek kihirdetésre a titkos határozatban előírányzott jogszabályok – többek közt a 6/1959. (II. 12.) FM-rendelet a termelőszövetkezetek gépi munkadíjkedvezményéről, a 13/1959. (III. 15.) *Kormányrendelet a mezőgazdasági termelőszövetkezetekben lépő termelők szerződéses kötelezettségének rendezéséről*, vagy az 1010/1959. (IV. 3.) *Kormányhatározat a termelőszövetkezeti mozgalom fejlődésével kapcsolatos szakember-átcsoportosítás egyes kérdéseiről*.



7. ábra. A Magyar Közlönyben kihirdetett agrárpolitikai jogszabályok altópikok szerint, évenként (N=445)

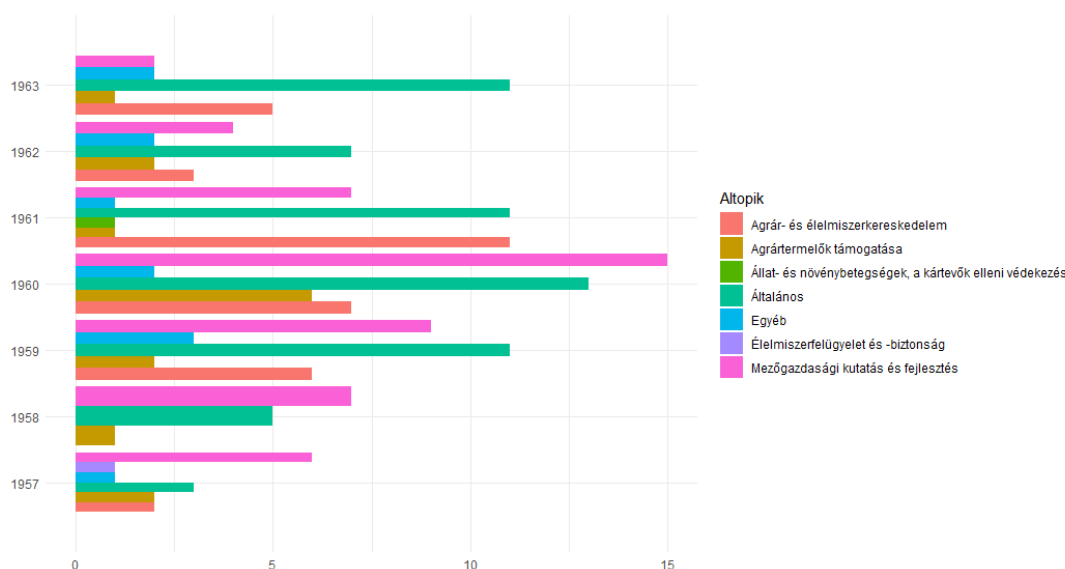


8. ábra. A Minisztertanács titkos agrárpolitikai hálózatai altópikok szerint, évenként (N=215)

Ha a Politikai Bizottság napirendi pontjainak altémakörök szerinti alakulását is meg-  
nézzük, sajátos fáziskésést figyelhetünk meg egyes nagy elemszámú topikok, így  
elsősorban az agrárkereskedelem témakör esetében. Az egyes konkrét előterjeszté-  
seket megvizsgálva az adatokból a Politikai Bizottság speciális szerepét, ellenőrző,  
nyomon követő funkcióját mutathatjuk ki. A bizottság által kezdeményezett intézke-  
dések mellett (pl. 1957. december 17. *Határozati javaslat a mezőgazdasági felvásárlási  
rendszer további javítására*; 1959. június 30. *Javaslat az 1959. évi bortermés elhelyezése*

érdekében szükséges intézkedésekre) nagy számban jelennek meg az ülések napirendjén olyan jelentések, beszámolók, amelyek a korábban jogszabályban vagy más utasításban elrendelt tevékenységek végrehajtásáról, a szabályozott folyamatok állásáról tájékoztatták a testületet (többek között 1960. szeptember 20. *Fehér Lajos tájékoztatója a burgonya felvásárlásával kapcsolatos intézkedésekről*; 1960. november 22. *Fehér Lajos tájékoztatója az őszi mezőgazdasági munkák és a felvásárlás helyzetéről*; 1961. október 17. *Jelentés a felvásárlás első háromnegyedévi tapasztalatairól*).

Míg az agrárkereskedelem esetében a Politikai Bizottság ellenőrző, „monitorozó” jellege tűnik erősebbnek, addig például a mezőgazdasági kutatás és fejlesztés területén egyértelműen kezdeményező szerepben van a testület – ez talán a terület nagyobb fokú generalitásával magyarázható. A fejlesztési irányvonalak általános („politikai szintű”) meghatározása, illetve az ezzel kapcsolatos javaslatok jóváhagyása egyértelműen a bizottság feladatkörébe tartozott.



9. ábra. Az MSZMP Politikai Bizottsága üléseinek agrárpolitikai napirendjei altópikok szerint, évenként (N=173)

#### 4.3. Szövegbányászati vizsgálat

Az előzőekben bemutatott témaalapú osztályozás mellett, amikor is az egyes dokumentumokat kézi kódolás segítségével klasszifikáltuk, szövegeinket szövegbányászati módszerekkel is vizsgálat alá vettük. Ahhoz, hogy egy szöveghalmazt közvetlenül kvantitatív módon elemezzünk, szükséges annak adattá alakítása. Ezt követhetik a különböző szövegbányászati vizsgálatok, amelyek segítségével kvantitatív tartalomelemzést végezhetünk, megvizsgálva például a szövegekben előforduló különböző szavak, illetve témák gyakoriságát vagy együttes előfordulását. Mindezek révén lehetőségünk nyílt a szöveghalmaz belső struktúrájának feltárására.<sup>40</sup>

<sup>40</sup> Németh, Katona és Kmetty, „Az automatizált szövegelemzés,” 55.

A következőkben *Natural Language Processing (NLP)* azaz a természetes nyelvfeldolgozás során alkalmazott egyik eljárást, a topikmodellezést hívtuk segítségül a mezőgazdasággal kapcsolatos szövegeink elemzéséhez. A természetes nyelvfeldolgozás során az alacsony mértékben strukturált szöveghalmazunkat kisebb egységekre bontottuk, és kapcsolatokat kerestünk a szöveg különböző elemei között. Ezzel az egyes szövegektől eltávolodva olyan ismeretek birtokába jutottunk, amelyek a szövegek egymás utáni olvasásával, illetve kvalitatív elemzéssel nem állnának rendelkezésünkre.<sup>41</sup>

Az elemzésbe a *Magyar Közlönyben* megjelent, a vizsgált korszakra vonatkozó és a CAP mezőgazdasági főtopikjába tartozó 445 darab jogszabály szövegét vontuk be. A szövegeket egyesével, jogszabályonként külön PDF-ben mentettük le az Optijus adatbázisából, majd ezeket az *ABBYY FineReader 14* optikai karakterfelismerő szoftver segítségével konvertáltuk TXT-formátumúvá, ügyelve arra, hogy a szövegek karakterkódolása UTF-8 legyen. Az *ABBYY FineReader* automatikusan állapítja meg a kódlapot, de manuálisan átállíthatjuk azt a kívánt kódtípusra. Erre azért szükséges figyelni, mert az UTF-8 kódolás egymagában képes az összes nyelv összes karakterét ábrázolni, így nem ütközhetünk abba a problémába, hogy mivel a program által automatikusan kiválasztott karakterkódolás nem képes a szövegben lévő egyes karakterek megjelenítésére, értelmetlen betűkapcsolatokat hozzon létre értelmes szöveg helyett.<sup>42</sup>

Az így keletkezett TXT-eket egységes struktúrájú elnevezéssel láttuk el, ami tartalmazta az egyes jogszabályok CAP-kódolt metaadattáblában szereplő azonosítóját. Mindez a későbbi visszakereshetőség szempontjából nagyon fontos lépés. A következőkben ezen szövegeinket kellett adatokká alakítanunk, amihez az *R* szoftverkörnyezetet használtuk.<sup>43</sup>

A szövegbányászati elemzések egyik igen fontos eleme az „előfeldolgozás” (*preprocessing*), ami több lépésből áll; ezek kiválasztása és elvégzésük sorrendje nagyrészt kutatói döntésen alapul. Mivel az előfeldolgozás folyamata során részben felesleges információktól szabadítjuk meg a szöveget, könnyen belátható, hogy az egyes lépéseket gondosan és a kutatási kérdésünknek megfelelően kell megterveznünk. Az előfeldolgozás első lépését még nem *R*-ben végeztük. A szöveget a *magyarlanc* morfológiai elemző segítségével szótöveztük, és meghatároztuk a benne szereplő szavak szófaját.<sup>44</sup> A szótövezés a statisztikai elemzések elvégzésének előfeltétele. A magyar nyelvben

<sup>41</sup> Bolonyai Flóra és Sebők Miklós, „Kvantitatív szövegelemzés és szövegbányászat,” in *Empirikus jogi kutatások: Paradigmák, módszertan, alkalmazási területek*, szerk. Jakab András és Sebők Miklós (Budapest: Osiris Kiadó, 2020), 361–380, 364.

<sup>42</sup> Hozzáférés: 2021.02.10, <http://www.cs.bme.hu/~egmont/utf8/>.

<sup>43</sup> Hozzáférés: 2021.02.10, <https://www.r-project.org/>. Az *R*-ben végzett szövegbányászati elemzések támogatására számos leírás érhető el, bár ezek használatát némiképp nehezíti, hogy gyakorlati példák angol nyelvű korpuszokra vonatkoznak, így a magyar nyelvű szövegek speciális elemzési kihívásaira nem reagálnak. A magyar nyelvű szövegek *R*-ben történő itt bemutatott elemzéséhez nagy segítségünkre voltak Labádi Gergely publikációi (hozzáférés: 2021.02.11, <http://labadiger.gy.github.io/>). Elemzésünk során az alábbi *R*-csomagokat használtuk: *tm*, *NLP*, *topicmodels*, *quanteda*, *tidytext*, *tidyr*, *ggplot*, *RTextTool*, *reshape2*, *dplyr*, *textmineR*, *LDavis*, *doParallel*, *scales*, *slam*.

<sup>44</sup> Hozzáférés: 2021.02.10, <https://github.com/zsibritajanos/magyarlanc>; János Zsibrita, Veronika Vincze and Richárd Farkas, „magyarlanc: A Toolkit for Morphological and Dependency Parsing of Hungarian,” in *International Conference Recent Advances in Natural Language Processing'2013: Proceedings*, eds. Ruslan Mitkov, Galia Angelova and Kalina Bontcheva (Shoumen: IN-COMA Ltd., 2013), 763–771.

a mondatokban a szavakhoz végződéses járulnak, illetve a szavaknak vannak elő- és utótagjai. Ezek eltávolításával (lemmatizálás) megkapjuk a szavak szótári alakját, azaz egységes alakra hozzuk az egy-többől különböző módon képzett vagy ragozott szavakat. A szófaj-meghatározás a *magyarlanc* futtatása során automatikusan megtörténik, így ha szükséges, később dönthetünk arról, ha az elemzésbe csak bizonyos szófajú szavakat szeretnénk bevonni. A természetes nyelvi feldolgozásban a szófaji egyértelműsítésnek gyakran fontos szerepe van, hiszen az ugyanolyan alakú szavak több jelentéssel is bírhatnak, ami egy adott aktuális kontextusban kulcsfontosságú lehet az elemzés szempontjából. A következő lépésben a korpuszt „stopszavaztuk”, azaz eltávolítottuk azokat a kifejezéseket, amelyek nem rendelkeznek tartalmi jelentéssel, ezek az úgynevezett funkciószavak (*function words*), például a kötőszavak.<sup>45</sup> Ehhez az *R* program *tm* programcsomagjában lévő, beépített magyar nyelvű stopszószótárát használtuk, amit a korpuszra jellemző stopszólistával egészítettük ki.<sup>46</sup> Az eddigiek eredményeként a korpuszunkban tehát a lemmatizált tartalomsszavak (*content words*) maradtak. Az előfeldolgozás következő lépése azon kifejezések kizárása az elemzésből, amelyek nagyon kevés, vagy nagyon sok dokumentumban szerepeltek. Például említhetjük a törvény, rendelet stb. szavakat, amelyek korpuszunkban értelemszerűen nagyon gyakoriak, de az elemzésbe nem tartottuk érdemesnek bevonni őket. Ehhez szógyakorisági listát készítettünk, ami a korpusz összes szavának gyakoriságát mutatta meg. Emellett készítettünk egy listát arról, hogy egy-egy adott kifejezés hány dokumentumban fordul elő. Végül az alsó eltávolítási küszöbértéket 0,1-ben a felsőt 0,4-ben határoztuk meg.<sup>47</sup>

Ugyancsak az előkészítő műveletek közé tartozik a névelem-felismerés és a szignifikáns *n*-gramok megkeresése. Ezek olyan több szóból álló kifejezések, amelyeket az elemzés során együtt szeretnénk kezelni. Ugyancsak eltávolítottuk a korpuszból a központosítást és egyéb speciális karaktereket, a letöltés következtében a szövegekbe került url-címeket, elvégeztük a szövegek kisbetűsítését. Az előkészítő lépések eredményeként korpuszunk 600797 tokent tartalmazott.

A korpuszépítés során a szövegek adattá konvertálására egy dokumentumkifejezőmátrixot (*document-term matrix, DTM*) hoztunk létre, amelyben minden dokumentum egy megfigyelés, azaz sor, és a dokumentumban szereplő minden kifejezés egy oszlop.<sup>48</sup>

A dokumentum kifejezőmátrix létrehozásával tulajdonképpen adattá alakítottuk szövegeinket. Ezután ugyancsak *R*-ben illesztettük az *LDA*-t, hogy ennek segítségével feltárjuk a mezőgazdasággal kapcsolatos jogszabályszovegek látens belső struktúráját.<sup>49</sup> Bár a topikok számának meghatározása kutatói döntésen alapul, és a modell

<sup>45</sup> Tikk, *Szövegbányászat*, 35; Burtejin, „Csoportosítás (klaszterezés),” 90.

<sup>46</sup> Hozzáférés: 2021.02.10, <https://cran.r-project.org/web/packages/tm/index.html>.

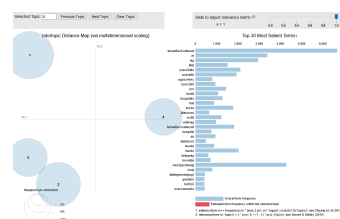
<sup>47</sup> Az előfeldolgozott szövegeket lásd a tanulmány mellékleteként (1030\_data.zip): <http://ojs.elte.hu/digitalisbolcseszett/article/view/1030/1808>. A szöveggörpuzs, a stopszavak, illetve a metaadatok elérhetők a GitHubon: [https://github.com/ringorsolya/lda\\_historical\\_sources\\_hun](https://github.com/ringorsolya/lda_historical_sources_hun).

<sup>48</sup> Hozzáférés: 2021.02.10, <https://www.rdocumentation.org/packages/tm/versions/0.7-7/topics/TermDocumentMatrix>.

<sup>49</sup> Hozzáférés: 2021.02.10, <https://www.rdocumentation.org/packages/lda/versions/1.4.2/topics/lda.collapsed.gibbs.sampler>.

futtatása során bevett gyakorlat a különböző topikértékekkel való kísérletezés, miután elkészült az elemzés több lehetőségünk is van az elkészült modell kiértékelésére. A zavarodottságfüggvény (*perplexity*) például a topikok által reprezentált elméleti szóeloszlásokat hasonlítja össze a szavak tényleges eloszlásával a dokumentumokban. Értéke nem önmagában értelmezendő, hanem két modell összehasonlításában, ahol a legalacsonyabb zavarodottságértékkel rendelkező modellt tekintik a legjobbnak. Elemzésünk során két *LDA*-módszert alkalmaztunk, az *LDA Gibbs*-et és a *VEM*-et, a validálás során ezek különböző topikszámokra adott zavarodottságeredményeit vetettük össze egymással. Az elemzés során 3 és 10 közötti topikszámokkal kísérleteztünk. A modell illesztése után minden esetben kiírtuk az egyes topikokba tartozó leggyakoribb 30 szót és megvizsgáltuk, hogy mely topikszám esetén kapunk jól interpretálhatóan egymástól elkülönülő topikokat. Az elemzésünket az *LDavis* segítségével vizuálisan is megjelenítettük, ami nagymértékben segítette ezen eredmények értelmezését. Mindezzel négy egymástól jól elkülöníthető nagy témát sikerült azonosítanunk.

Hogy elemzésünket kvalitatív módon is kiértékeljük az *R* program segítségével azt is meghatároztuk, melyik jogszabály mely topikba tartozik inkább, majd ezt összevetettük a jogszabályok CAP-altopikba sorolásának eredményével is. Az összevetést interaktív HTML-es vizualizáció segíti.



A dinamikus HTML-tartalom megtekintéséhez kattintson a képre  
<http://ojs.elte.hu/digitalisbolcseszett/article/view/1030/1810>

Az első topikot elsődlegesen a mezőgazdasági termelés és termelési körülmények (a vetőmagellátás, szerződéses termelés, terményfelvásárlás) fogalmai határozzák meg. A második topik a domináns fogalmak alapján a termelőszövetkezeti működés, a harmadik topik a mezőgazdasági terményforgalom és ezen belül a jogszabályokban kiemelt figyelemmel övezett bortermelés és borkereskedelem, míg a negyedik az állat- és növény-egészségügyi védekezés témaköréhez kapcsolható.

Elemzésünk harmadik részében kísérletet tettünk a két eddig alkalmazott vizsgálati módszer, az előre megadott kategóriák mentén történő kézi osztályozás, valamint a szövegbányászat kombinálására. Célunk, hogy összefüggést találjunk a CAP-kódkönyvben meghatározott közpolitikai topikok, konkrétan az agrárpolitikai altopikok, valamint a topikmodell során kapott témák között.

A két eljárás összekapcsolása érdekében a jogszabályokat a fentiekben már jelzettek szerint hozzárendeltük a topikmodell-elemzés során azonosított négy topikhoz. Mivel a jogszabályok CAP-kódolása során a CAP-altopikba rendelt már elvégeztük, így az egyes jogszabályok esetében lehetségessé vált a CAP-altopik és a topikmodell szerinti besorolás együttes vizsgálata.

	Mezőgazdasági termelés	Termelőszövetkezeti működés	Mezőgazdasági terményforgalom	Állat- és növény-egészségügyi védekezés	Összesen
Általános	20,5%	9,0%	55,1%	15,4%	100,0%
Agrár- és élelmiszer-kereskedelem	35,4%	21,5%	20,8%	22,3%	100,0%
Agrártermelők támogatása	12,5%	16,7%	50,0%	20,8%	100,0%
Élelmiszer-felügyelet és -biztonság	16,3%	36,3%	13,8%	33,8%	100,0%
Állat- és növénybetegségek	20,9%	17,9%	20,9%	40,3%	100,0%
Halászat és horgászat	16,7%	25,0%	25,0%	33,3%	100,0%
Mezőgazdasági kutatás és fejlesztés	12,5%	12,5%	62,5%	12,5%	100,0%
Egyéb	18,2%	18,2%	31,8%	31,8%	100,0%
Összesen	22,9%	20,7%	30,1%	26,3%	100,0%

1. táblázat. CAP-altopik-besorolás és a topikmodell szerinti besorolás együttes vizsgálata

CAP-altopik	Elemszám	Százalék
Általános	78	17,5%
Agrár- és élelmiszer-kereskedelem	130	29,2%
Agrártermelők támogatása	48	10,8%
Élelmiszer-felügyelet és -biztonság	80	18,0%
Állat- és növénybetegségek	67	15,1%
Halászat és horgászat	12	2,7%
Mezőgazdasági kutatás és fejlesztés	8	1,8%
Egyéb	22	4,9%
Összesen	445	100,0%

2. táblázat. A vizsgált jogszabályok megoszlása CAP-altopikok szerint

LDA-topik	Elemszám	Százalék
1	102	22,9%
2	134	30,1%
3	92	20,7%
4	117	26,3%
Összesen	445	100,0%

3. táblázat. A vizsgált jogszabályok megoszlása a topikmodell egyes topikjai szerint

Látható, hogy míg a topikmodell-vizsgálat során négy, közel azonos elemszámú topikot kaptunk, addig a CAP-altopikok elemszáma nagyobb eltéréseket mutat. Kiemelkedő jelentőségű az „agrár- és élelmiszer-kereskedelem” altopik (a jogszabályok közel 30%-a tartozik ide), míg a „halászat és horgászat”, a „mezőgazdasági kutatás és fejlesztés”, valamint a nem egyértelműen besorolható elemeket tartalmazó „egyéb” altopikok együttesen sem teszik ki az összes vizsgált jogszabály 10%-át.

Már a két klaszterváltozó egyszerű keresztmegoszlása alapján is kirajzolódni látszik a CAP-altopikok és a topikmodell eredményei közötti összefüggés. Az „általános” elnevezésű altopikba sorolt jogszabályok 55%-a, az „agrártermelők támogatása” altopik jogszabályainak fele a második topikba került, hasonlóan a kis elemszámú „mezőgazdasági kutatás és fejlesztés” altopik elemeinek közel egyharmadához, míg az „állat- és növénybetegségek” altopik jogszabályainak 40%-a a negyedik topikba sorolódott.

Az „agrár- és élelmiszer-kereskedelem” altopik jogszabályainak 35,4%-a került az első topikba, míg az „élelmiszer-felügyelet és -biztonság” altopik esetében közel azonos, 33% illetve 36% a harmadik és a negyedik topikba sorolt jogszabályok aránya.

A két eljárás topikjainak eltérő elemszáma természetesen befolyással van a százalékos értékek alakulására, így a pontosabb eredmény miatt az adjusztált standardizált reziduálisok vizsgálatával cellánként is megkerestük a CAP-altopikok és a topikmodell topikjai közötti szignifikáns összefüggéseket. Az „általános” altopik jogszabályai a második topikba, az „agrár- és élelmiszer-kereskedelem” jogszabályai az első topikba sorolódtak a függetlenség esetén várható eloszláshoz képest szignifikánsan nagyobb arányban. Ezt az összefüggést mutatja az „agrártermelők támogatása” valamint a „mezőgazdasági kutatás és fejlesztés” altopik és a második topik, az „élelmiszer-felügyelet és -biztonság”, továbbá a harmadik topik, az „állat- és növénybetegségek” altopik és a negyedik topik is. A „halászat és horgászat” esetében nem találtunk elég erős összefüggést a topikmodell egyes topikjaival.

A magyarországi agrárpolitika történeti bemutatásánál ejtettünk szót a termelőszövetkezetek működési problémáiról, a „gyengén teljesítő” termelőszövetkezetek kérdéséről, ezek támogatásának agrárpolitikai vonatkozásairól. Ha az „agrártermelők támogatása” CAP-altopik és a topikmodell eredményeinek összefüggését nézzük, egyértelműen kirajzolódik az a tendencia, hogy vizsgált időszakunkban az agrártermelői támogatás elsődlegesen a tsz-ek működésének támogatásával volt azonos. Az „általános” altopik és a topikmodell „termelőszövetkezeti” topikjának erős kapcsolata jelzi az agrárpolitikai igazgatás fő irányát, az újonnan kialakított termelőegységek működésének, működési környezetének erőteljes szabályozását.

Kiemelendő, hogy a topikmodell-elemzés során „mezőgazdasági terményforgalom”-ként azonosított topik az „élelmiszer-felügyelet és -biztonság” CAP-altopikkal mutat



erős kapcsolatot, míg az „agrár- és élelmiszer-kereskedelem” altopik a „termelés és termelési körülmények” topikkal. A mezőgazdasági termékek forgalmazásánál a két eljárás összefüggése szerint elsősorban az élelmiszer-biztonsági, egészségügyi vonatkozások domináltak, míg a növénytermesztés általános szabályozása leginkább a termékek felvásárlásának, felvásárlási árának, a termelési körülményeknek, a termelőegységek azonosításának a kérdéskörével foglalkozott. Mindez egyértelműen összhangban van korábbi megállapításunkkal a mezőgazdasági termelési és termékkereskedelmi feltételrendszer kialakításának szükségességéről.

A nagy méretű szövegtörzsek elemzésére több lehetőség is rendelkezésünkre áll. Ilyen lehetőség a törzs különböző szakaszainak címkékkel való ellátása, amire tanulmányunkban a CAP-kódolás alkalmazását hoztuk példaként. Ennek során deduktív módon, előre meghatározott kategóriákat használtunk és kézi kódolást végeztünk. Az ilyen típusú osztályozás előnye lehet, hogy a kutatási kérdéstől függően mélyebb elemzést tesz lehetővé, ellenben jelentős emberierőforrás-igénye van, képzett kutatók szükségesek hozzá, és még így is számolni kell az inkonzisztencia lehetőségével. Az osztályozás végezhető gépi támogatással is, vagy felügyelt gépi tanuló algoritmusok segítségével. Ezek jóval kisebb emberi erő bevonását teszik szükségessé, magát az elemzést felgyorsítják, de jelentős előkészítő munkát igényelnek, például tanulótorzs létrehozását, eredmények validálását. Az ilyen előre meghatározott kategóriák előnye, hogy az egységesen kódolt törzsek segítségével lehetőséget ad a folyamatok hosszú időintervallumban történő vizsgálatára, hátrányuk ellenben, hogy kevésbé igazíthatóak egy-egy törzs egyedi jellemzőihez. Bár a CAP-kódrendszer nem minden esetben illeszthető könnyen a történeti kutatások szempontrendszeréhez, és az eltérő történelmi korszakok közpolitikájának tagolása, az egyes kategorizálandó elemek témakörökbe sorolása (különösen az altémakörök szintjén) is okozhat nehézségeket, alapvetően alkalmas arra, hogy a közpolitikai figyelem alakulásának történeti elemzéséhez felhasználjuk. Ezen altémakörök szintjén megjelenő kategorizálási nehézségekre jelenthet megoldást a dokumentumok klaszterezése.

Az esettanulmányunk során bemutatott felügyeletlen gépi tanulás körébe tartozó topikmodellezés ugyanis az adott törzs jellegzetességeiből indul ki és segít eligazodni azokban a törzsekben, amelyek tartalmáról olyan keveset tudunk, hogy máskülönben érdemi vizsgálatuk megkezdéséhez hosszas tanulmányozásukra lenne szükségünk. Emellett még az előkészítő lépésekkel együtt is lényegesen gyorsabb elemzést tesz lehetővé, mint a kézzel végzett osztályozás.

## 5. Összegzés és további kutatási irányok

Tanulmányunkban a magyarországi kollektivizálás második hullámát egy, a társadalomtudományban használatos módszer adaptálásának segítségével vizsgáltuk meg. Bemutattuk, hogy a nagy adatbázisok elemzése milyen nagyobb, általánosabb összefüggések levonására kínál lehetőségeket. Ugyanezt a kutatási kérdéskört kvantitatív szövegelemzési módszertannal is megvizsgáltuk, az egyik legelterjedtebb szöveg-bányászati módszer, az úgynevezett csoportosítás vagy klaszterezés, és azon belül a topikmodellezés alkalmazásával. Célunk annak illusztrálása volt, hogyan lehet új

elemzési eszközökkel, más tudományágak által alkalmazott modellek segítségével tovább mélyíteni egy történelmi folyamatról korábban megalkotott képet.

Egyetértve Z. Karvalics Lászlóval, úgy véljük, hogy önmagában sem a *big data* jelenség és annak a történelmi kutatásokra gyakorolt hatása, így a nagy digitális gyűjtemények létrejötte és folyamatos gyarapodása, sem azok elemzésének új módszerei nem hoznak ugyan paradigmátikus változást a történelmi megismerés módjában, viszont számtalan új módszertani lehetőségnek nyitnak utat.<sup>50</sup> Ezek az új módszerek azonban nem helyettesítik a hagyományos kutatási formákat és módszereket, inkább kiegészítik, támogatják azokat. A kvantitatív és kvalitatív források és módszertanok mellett a szövegelemzés kvantitatív és kvalitatív formái is kiegészítik egymást. De a választott módszertannak mindig a kutatói döntésen és a kutatási kérdéseken, valamint a rendelkezésre álló forrásokon kell alapulnia.

Az egyre gyarapodó digitalizált dokumentumállomány ellenére a szövegbányászati módszerek széles körű történettudományi alkalmazása előtt jelenleg nagy gátat jelent a jó minőségű, nagy méretű korpuszok és az egységesen metaadatolt adatbázisok hiánya. Mint rámutattunk, önmagában a digitális állományok megléte nem teszi azokat szövegbányászati vizsgálatra alkalmassá, az egyik legnagyobb leküzdendő akadályt a források szöveggé alakítása előtt az optikai karakterfelismerés hiányosságai jelentik. A papíron keletkezett dokumentumok sokszor csak hosszadalmas manuális tisztítás után válnak szövegbányászati módszerekkel elemezhetővé.

Ahogy arra korábbi tanulmányában Péter Róbert is rámutatott, olyan kifinomult kutatási eszközökre és szoftverekre is szükség van, amelyek úgy képesek megbirkózni hatalmas mennyiségű szöveggel és metaadattal, hogy az elemzésnél a bölcsészettudományi kutatásban hangsúlyos részletek sem vesznek el.<sup>51</sup> Ehhez nagy segítséget jelenthet a különböző tudományterületek összefogása, az interdiszciplináris kutatócsoportok működése. Mindamellet azon a véleményen vagyunk, hogy megalapozott történettudományi kutatás digitális bölcsészeti módszerekkel csak úgy végezhető, ha megfelelő képzés is járul hozzá, azaz a módszer iránt érdeklődő történészek számára lehetőség nyílik a bölcsészettudományi kutatások kérdésfeltevéseihez és módszertanához igazodó digitális bölcsészeti képzettség megszerzésére, ami egyben azt is lehetővé tenné, hogy szakmai diskurzus induljon e módszerek használatáról, amelyek történettudományi kérdésfeltevés és releváns források ismerete mellett azonban kiválóan egészíthetik ki a hagyományos kutatási módszerek alkalmazását.

<sup>50</sup> Z. Karvalics László, „A nagy adat és digitális történelem: Egy izgalmas házasság múltja, jelene és jövője,” *Magyar Tudomány* 179, 5. sz. (2018): 670–671, <https://doi.org/10.1556/2065.179.2018.5.10>.

<sup>51</sup> Péter, „A Big Data kihívás,” 1328–1329.

## **Agricultural Policy Challenges and Legislation in the Early Kádár Era: Text Analysis and Text Mining of Historical Sources**

In our study, we examine the collectivization in Hungary by adapting a method used in political science and analyse large databases. In addition we examine the same question using text mining method. The aim of our paper is to illustrate the possibilities and limitations of the analysis of digital historical collections with new research methods. Our analysis examines the agricultural legislation of the early Kádár era. In addition to the classification, we also examine our texts using text mining methods, and in the third part of our analysis we try to combine the two methods – to find a correlation between agricultural policy subtopics and topics obtained in the topic model.

Keywords:

digital humanities, history, classification, text mining, topic modelling



<MŰHELY>



**Horváth Péter**

*ELTE Digitális Bölcsészet Központ*

horvath.peeteer@gmail.com

## **A vershangzás jellemzőinek automatikus feltárása József Attila verseiben\***

A tanulmány József Attila versei alapján mutatja be a vershangzáshoz kapcsolódó tulajdonságok automatikus feltárásának egy módszerét. A tanulmány első fele ismerteti a vershangzás jellemzőinek automatikus annotálására létrehozott *hunpoem\_analyzer-TEI* program funkcióit, valamint az annotációs folyamat főbb lépéseit. A tanulmány második fele különböző, az elemzett korpuszból kinyert, a versek lexikai tulajdonságaihoz és a vershangzás jellemzőihez kapcsolódó gyakorisági listákat mutat be.

Kulcsszavak:

automatikus verselemzés, korpusznyelvészet, vershangzás, rím, ritmus, József Attila, *hunpoem\_analyzer-TEI*, *e-magyar*



### **1. Bevezetés**

Számítógépes, automatikus módszerek használata lehetővé teszi nagy mennyiségű vers elemzését, ami manuálisan nem, vagy csak nagyon hosszú idő alatt lenne elvégezhető. Az elemezhető versek mennyiségének ez az ugrásszerű növekedése újfajta tudományos kérdések megválaszolását teszi lehetővé. Teljes életművekkel, korszakokkal vagy akár költészeti hagyományokkal kapcsolatban fogalmazhatunk meg kvantitatív alapokon nyugvó állításokat. Ennek a tanulmánynak ugyanakkor nem célja József Attila költészetével kapcsolatban irodalomtudományos megállapításokra jutni. Az alábbiakban pusztán azt mutatom be, hogy milyen típusú jellemzőket nyerhetünk ki egy verset tartalmazó korpuszból számítógépes eszközök segítségével. A József Attila-versek hangzásjellemzőinek az elemzésére használt *hunpoem\_analyzer-TEI* elnevezésű programot a készülő ELTE Verskorpusz számára írtam. A program jelenleg is fejlesztés alatt áll, vagyis a tanulmányban bemutatott funkciói nem tekinthetők véglegesnek.

A tanulmány második részében bemutatok néhány nemzetközi példát versek automatikus elemzésére, a harmadik részben pedig ismertetem a vershangzás jellemzőinek

\* Köszönöm a tanulmány két névtelen lektorának a részletes javaslatokat. Javasataik nem csupán ennek a tanulmánynak a megírásában segítettek sokat, hanem a vershangzást elemző program továbbfejlesztésében is nagyon hasznosak lesznek.

az elemzésére használt *hunpoem\_analyzer-TEI* program funkcióit. A negyedik részben röviden áttekintem a József Attila-korpusz létrehozásának és automatikus elemzésének főbb lépéseit. A tanulmány ötödik részében néhány, a versek szókészletére, illetve grammatikai tulajdonságaira vonatkozó jellemzőt mutatok be, melyek az MTA Nyelvtudományi Intézetében fejlesztett *e-magyar* programnak a verseken való lefutásával váltak kinyerhetővé. A hatodik részben a korpusznak a *hunpoem\_analyzer-TEI* program futtatása révén kinyert, vershangzáshoz kapcsolódó jellemzőit ismertetem, az utolsó, hetedik részben pedig röviden összegzem a tanulmányban bemutatott módszert.

## 2. Néhány példa a vershangzás jellemzőinek automatikus elemzésére

Grammatikai tulajdonságok automatikus elemzése Magyarországon is jól ismert eljárás. Például a magyar nyelvészeti kutatásokban széles körben használt Magyar Nemzeti Szövegtár kereshető formában tartalmazza a szavak automatikus elemzés útján létrehozott grammatikai annotációit.<sup>1</sup> A vershangzáshoz kapcsolódó jellemzők automatikus elemzése azonban kevésbé gyakori eljárás, így az itt ismertetendő módszer előtt érdemes néhány, a vershangzás automatikus elemzését megcélzó nemzetközi példát is röviden bemutatni. Az alábbiakban szereplő eszközök, korpuszok és kutatások az utóbbi tizenöt évhez kapcsolódnak. Hangsúlyozandó, hogy az áttekintés ezen időszakra vonatkozóan sem kimerítő jellegű.

A vershangzáshoz kapcsolódó különböző jellemzők közül kifejezetten intenzíven foglalkoznak a ritmus automatikus elemzésével. A Charles O. Hartman által fejlesztett *Scandroid* nevű program például angol nyelvű versek ritmusát elemzi.<sup>2</sup> A program az elemzés során megkülönbözteti a vers szavainak hangsúlyos és hangsúlytalan szótagjait, majd ez alapján eldönti, hogy a vers jambikusnak vagy anapesztikusnak tekinthető-e inkább, és megállapítja az egyes verslábakat.<sup>3</sup> A hangsúlyos és hangsúlytalan szótagok megkülönböztetése részben általános szabályok alapján, részben pedig a programba beépített lexikon alapján történik. A *Scandroid* program módosított, nagy adatmennyiségen lefuttatható változatát használja Chris Tanasescu, Bryan Paget és Diana Inkpen, akiknek a kutatása angol nyelvű versek ritmus és rím alapján történő automatikus osztályozására irányul.<sup>4</sup>

Az úgyszintén angol nyelvű versek elemzésére fejlesztett *ZuScansion* nevű eszköz a *Scandroid*tól eltérően a jambikus és anapesztikus metrumok mellett egyéb metrumok

<sup>1</sup> Csaba Oravecz, Tamás Váradi and Bálint Sass, „The Hungarian Gigaword Corpus,” in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014)*, eds., Nicoletta Calzolari, Khalid Choukri, Thierry Declerck et al. (Reykjavik: European Language Resources Association [ELRA], 2014), 1719–1723.

<sup>2</sup> Charles O. Hartman, *The Scandroid. Version 1.1. [User guide]* (2005), hozzáférés: 2020.01.19, <http://charlesohartman.com/verse/scandroid/ScandroidManual.pdf>.

<sup>3</sup> Az angol (és más indoeurópai nyelvek) versritmusa nem a rövid és hosszú szótagok, hanem a hangsúlyos és hangsúlytalan szótagok váltakozására épül.

<sup>4</sup> Chris Tanasescu, Bryan Paget and Diana Inkpen, „Automatic Classification of Poetry by Meter and Rhyme,” in *AAAI Publications: The Twenty-Ninth International Flairs Conference (2016)*, eds., Zdravko Markov and Ingrid Russell, hozzáférés: 2020.01.19, <https://www.aaai.org/ocs/index.php/FLAIRS/FLAIRS16/paper/view/12923/12883>.



elkülönítésére is képes.<sup>5</sup> A metrum megállapításának alapját az elsődleges hangsúlyú, a másodlagos hangsúlyú és a hangsúlytalan szótagok automatikus elkülönítése adja. A program a szótagok hangsúlyait egyrészt két kiejtésszótár alapján, másrészt a lexikai hangsúlyt adott esetben felülíró, a szavak szófajához kapcsolódó prozódiai szabályok alapján állapítja meg. Amennyiben egy szó nem szerepel a program által használt szótárakban, akkor az algoritmus az írásmódja alapján az elemzendő szóra leginkább hasonlító szótári szó hangsúlyviszonyai alapján elemzi azt. A program a versre általánosan jellemző metrumot oly módon állapítja meg, hogy kiszámítja az egyes szótaghelyekre eső szótagok átlagos hangsúlyértékét, majd az ebből absztrahált ritmusképletnek több alternatív, verslábakra tagolt elemzését is megadja, amelyekből egy pontozási rendszer alapján választja ki a legmegfelelőbbet (az algoritmus azonos szótagszámú sorokon működtethető).

Az *AnalysePoem* program angol nyelvű versek ritmusának és rímképletének az elemzésére lett fejlesztve.<sup>6</sup> A ritmus elemzésének első lépéseként a program a beépített és a felhasználó által bővíthető lexikonja alapján, a *ZeuScansion*höz hasonlóan megkülönböztet erősen és gyengén hangsúlyos, valamint hangsúlytalan szótagokat. Amennyiben egy szó nem szerepel a lexikonban, az algoritmus általában meg tud adni a szóra egy valószínűsíthető elemzést. A szótagok hangsúlyai alapján a program hat domináns metrumot képes megállapítani. A domináns metrum megállapításához a program számos tesztet végez el egymást követően, amelynek kimeneteként egy megbízhatósági értéket (*confidence number*) rendel a megállapított metrumhoz. Minél nagyobb a megbízhatósági érték, annál szabályosabb a ritmus, egy bizonyos megbízhatósági érték alatt a vers nem tekinthető az adott metrumhoz tartozónak. Az *AnalysePoem* a versek rímképletét is képes megadni, aminek előfeltétele a ritmus elemzése. A program a rímképletek megfelelő elemzéséhez egy folyamatosan bővülő adatbázist használ, amely a korábban már elemzett, egymással rímelő szavakat tartalmazza.

Justine Kao és Dan Jurafsky kutatása professzionális és amatőr amerikai versek különbségeit vizsgálja egy amatőr és egy professzionális verseket tartalmazó, a kutatás számára létrehozott korpusz különböző nyelvi tulajdonságainak kvantitatív összevetésével.<sup>7</sup> A szókincs elemzése mellett a szerzők a David Kaplan által fejlesztett *Poetry-Analyzert* használva automatikusan elemzik a versekben szereplő alliterációkat (egymást követő, azonos mássalhangzóval kezdődő szavak), asszonáncokat (ugyanazon magánhangzók ismétlődése), konzonáncokat (ugyanazon mássalhangzók ismétlődése) és a sorvégi rímpárokat. A rímpárok elkülönítése során a program tiszta rímeket (*perfect rhyme*) és nem tiszta rímeket (*slant rhyme*) különböztet meg. Tanasescu, Paget és Inkpen már idézett, a ritmus mellett a sorvégi rímek automatikus elemzésére is kiterjedő kutatása a rímek két csoportja, a tiszta rím (*perfect rhyme*) és a Kao és Jurafsky kutatásánál tágabban értelmezett nem tiszta rím (*slant rhyme*) mellett egy

<sup>5</sup> Manex Agirrezabal, Aitzol Astigarraga, Bertol Arrieta and Mans Hulden, „ZeuScansion: A Tool for Scansion of English Poetry,” *Journal of Language Modelling* 4, 1. sz. (2016): 3–28.

<sup>6</sup> Marc R. Plamondon, „Virtual Verse Analysis: Analysing Patterns in Poetry,” *Literary and Linguistic Computing* 21, 1. sz. (2006): 127–141, <https://doi.org/10.1093/llc/fql011>.

<sup>7</sup> Justine Kao and Dan Jurafsky, „A Computational Analysis of Style, Affect, and Imagery in Contemporary Poetry,” in *Proceedings of the NAACL-HLT 2012 Workshop on Computational Linguistics for Literature*, eds., David Elson, Anna Kazantseva, Rada Mihalcea et al. (Montréal: Association for Computational Linguistics, 2012), 8–17.

harmadik csoportot, az *eye rhyme* kategóriát is bevezeti. Ide azok az esetek tartoznak, amikor a szavak az írásmódjuk miatt tiszta rímnek tűnnek, de valójában nem rímelnek (pl. *slaughter-laughter*).

Nem angol nyelvű versek automatikus elemzésére is találhatunk példákat. A versek szerkezeti elemei mellett a ritmus annotálására törekszik például a 16. és 17. századi spanyol szonettek tartalmazó Corpus de Sonetos del Siglo de Oro (Corpus of Spanish Golden-Age Sonnets).<sup>8</sup> A korpusban egy erre fejlesztett programmal automatikusan annotálják a verssorokban a hangsúlyos és hangsúlytalan szótagokat, a nem egyértelmű ritmusú, többféleképpen is annotálható verssorokat pedig manuálisan ellenőrzik. Érdemes megemlíteni a Cseh Tudományos Akadémia által fejlesztett Cseh verskorpuszt (Korpus českého verše) is, amely közel 80 000 annotált verset tartalmaz a 19. századból és a 20. század elejéről.<sup>9</sup> A korpusz a szavak lemmája, szófaja, morfológiai és fonológiai jellemzői mellett a ritmusra és a rímekre vonatkozó automatikusan létrehozott annotációkat is tartalmaz.<sup>10</sup> A korpusz honlapján számos lekérdező eszközből választhatunk.<sup>11</sup>

### 3. A *hunpoem\_analyzer-TEI* program funkciói

A József Attila-versek hangzásjellemzőinek automatikus elemzésére az általam írt *hunpoem\_analyzer-TEI* programot használtam, amely a készülő ELTE Verskorpusz hangzásjellemzőinek az annotálására lett fejlesztve.<sup>12</sup> A Python nyelvben írt program bemenetét a versek szerkezeti egységeinek, azaz a címnek, a versszakoknak és a soroknak a szövegközi (*inline*) annotációit tartalmazó TEI XML-fájlok adják.<sup>13</sup> A program két fő modulból tevődik össze. Az első modul az MTA Nyelvtudományi Intézetében fejlesztett *e-magyar* program *emtsv* nevű változatát futtatja, és a szavaknak az *e-magyar* TSV-kimenetében szereplő grammatikai elemzésével, azaz a szavak lemmá-

<sup>8</sup> Borja Navarro-Colorado, „A Computational Linguistic Approach to Spanish Golden Age Sonnets: Metrical and Semantic Aspects,” in *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, eds., Anna Feldman, Anna Kazantseva, Stan Szpakowicz et al. (Denver: Association for Computational Linguistics, 2015), 105–113, <https://doi.org/10.3115/v1/w15-0712>; Borja Navarro-Colorado, Marí Ribes Lafoz and Noelia Sánchez, „Metrical Annotation of a Large Corpus of Spanish Sonnets: Representation, Scansion and Evaluation,” in *Proceedings of the 10th Edition of the Language Resources and Evaluation Conference (LREC 2016)*, eds., Nicoletta Calzolari, Khalid Choukri, Thierry Declerck et al. (Portorož: ELRA, 2016), 4360–4364.

<sup>9</sup> Petr Plecháč and Robert Kolár, „The Corpus of Czech Verse,” *Studia Metrica et Poetica* 2, 1. sz. (2015): 107–118.

<sup>10</sup> Robert Ibrahim and Petr Plecháč, „Toward Automatic Analysis of Czech Verse,” in *Formal Methods in Poetics*, eds., Barry P Scherr, James Baily and Evgeny V. Kazartsev (Lüdenscheid: RAM, 2011), 295–305.

<sup>11</sup> Hozzáférés: 2020.01.19, [http://versologie.cz/v2/web\\_content/tools.php?lang=en](http://versologie.cz/v2/web_content/tools.php?lang=en).

<sup>12</sup> Köszönöm Indig Baláznak a kód átnézését, és a kevésbé szerencsés kódolási megoldások javítását. Emellett köszönöm neki az *e-magyar* futtatásában nyújtott segítséget.

<sup>13</sup> TEI Consortium, *TEI P5: Guidelines for Electronic Text Encoding and Interchange. Version 3.5.0.* (2019), hozzáférés: 2020.01.19. <https://tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf>.

jával (szótári alakjával), szófajával és morfoszintaktikai tulajdonságaival kiegészíti a TEI XML-fájlokat.<sup>14</sup>

A második modul végzi el a vershangzás jellemzőinek elemzését, amelynek során az elemzett jellemzők is bekerülnek annotációként a TEI XML-fájlokba. A program automatikusan megadja a szavak főbb fonológiai tulajdonságait, a versek versszakainak rímképletét, a rímpárokat alkotó szavakat, a verssorok ritmusát, valamint az alliterációkat alkotó szavakat. A szavak fonológiai tulajdonságainak elemzése a szótagszám, a hangrend (magas, mély vagy vegyes),<sup>15</sup> valamint a szó fonológiai szerkezetének a megadását jelenti. A fonológiai szerkezet elemzése során minden szó kap egy C, V, B, F, 1 és 2 karakterekből álló karaktersort, amelyben a karakterek a szó hangjainak néhány fontosabb fonológiai tulajdonságát jelölik. Ezek a következők: C – mássalhangzó, V – magánhangzó, B – hátul képzett magánhangzó, F – elöl képzett magánhangzó, 1 – rövid magánhangzó, 2 – hosszú magánhangzó (pl. „szerszámaival” – C, VF1, C, C, VB2, C, VB1, VF1, C, VB1, C).

A rímképlet elemzése során a program minden versszak esetében megadja annak rímképletét a megszokott módon, az ábécé betűiből álló karaktersorral (pl. aaaa, aabbcc, abcb).<sup>16</sup> A program azokat a sorvégeket tekinti egymással rímelőnek, amelyekben az utolsó szótag magánhangzója a hosszúságot nem számítva megegyezik, valamint megegyezik az utolsó előtti szótagok hosszúsága. A rímelés e szabályának alkalmazásában a fő szempont az volt, hogy a szabály ne legyen túl szűkös, de ne is generáljon túl.<sup>17</sup> Mindkét eset ugyanis ahhoz vezet, hogy a konzisztensen, azaz azonos rímképletű versszakokkal is leelemezhető verseket a program nagyobb eséllyel kezelné inkonzisztens módon, vagyis a túl specifikus vagy túl általános szabály alkalmazása miatt bizonyos versszakokat a többihez képest eltérő rímképlettel annotálna. Természetesen a későbbiekben a rímelés automatikus elemzése finomítható. Valószínűleg ki lehetne dolgozni olyan algoritmust, amely a konzisztens elemzés elvének a megtartása mellett az előző részben bemutatott példákhoz hasonlóan elkülöníti egymástól a tiszta

<sup>14</sup> Váradi Tamás, Simon Eszter, Sass Bálint, Geröcs Mátyás, Mittelholtz Iván, Novák Attila, Indig Balázs, Prószyk Gábor és Vincze Veronika, „Az e-magyar digitális nyelvfeldolgozó rendszer,” in *XIII. Magyar Számítógépes Nyelvészeti Konferencia*, szerk., Vincze Veronika (Szeged: Szegedi Tudományegyetem, Informatikai Intézet, 2017), 49–60; Mittelholtz Iván, „emToken: Unicode-képes tokenizáló magyar nyelvre,” in *XIII. Magyar Számítógépes Nyelvészeti Konferencia*, szerk. Vincze, 61–69; Novák Attila, Rebrus Péter, Ludányi Zsófia, „Az emMorph morfológiai elemző annotációs formalizmusa,” in *XIII. Magyar Számítógépes Nyelvészeti Konferencia*, szerk. Vincze, 70–78; Indig Balázs, Sass Bálint, Simon Eszter, Mittelholtz Iván, Kundráth Péter és Vadász Noémi, „emtsv – Egy formátum mind felett,” in *XV. Magyar Számítógépes Nyelvészeti Konferencia*, szerk., Berend Gábor, Gosztolya Gábor és Vincze Veronika (Szeged: Szegedi Tudományegyetem TTIK, Informatikai Intézet, 2019), 235–247.

<sup>15</sup> A program a hangrendnek a hagyományos, közoktatásban alkalmazott megközelítését használja azzal a módosítással, hogy az összetett szavak esetében a hangrend megállapítása nem az utolsó tag, hanem a teljes szó alapján történik, hiszen a hangrend annotálásának a célja esetleges poétikai következtetések elősegítése, nem pedig a toldalékok magánhangzóinak az előrejelzése (amire egyébként a közoktatásban alkalmazott hagyományos megközelítés nem is igazán alkalmas).

<sup>16</sup> Az elemzőprogram egy lehetséges továbbfejlesztésének iránya, hogy ne csak a versszakon belüli, hanem az interstrofikus, versszakon túlnyúló rímelést is elemezni tudja. Például a József Attila költszetében is megjelenő szonettformára jellemző ez a rímelési mód.

<sup>17</sup> Egy, az alkalmazottnál még általánosabb szabály lenne például, ha a program nem venné figyelembe a sor utolsó előtti szótagjának a hosszúságát, vagyis minden azonos magánhangzóra végződő sorvéget egymással rímelőnek tekintene.

és nem tiszta rímeket (azaz az asszonáncokat), adott esetben a nem tiszta rímeket is több csoportba sorolva.<sup>18</sup>

A rímképlet mellett a program megadja az egymással rímpárt alkotó, azaz hívó- és felelőrim viszonyban lévő szavak listáját, rímpáronként elkülönítve azokat. A program jelenlegi verziója csak versszakon belül elemzi rímpárokat. Egy rímpár szavai között maximum négy sor lehet, például egy hat soros abbcca rímképletű versszakban az első és az utolsó sor rímhelyzetben lévő szavait rímpárként azonosítja a program, de például egy abbccca rímképletű hétsoros versszak első és utolsó sorának rímhelyzetben lévő szavait már nem elemzi rímpárként. Egy rímhelyzetben lévő szó két rímpárnak is a része lehet, az első rímpárban felelő-, a másodikban hívórímként. Például egy négysoros bokorrím, azaz egy aaaa rímképletű versszak második sorának a rímhelyzetben lévő szava felelőrímként az első sor, hívórímként pedig a harmadik sor sorvégi szavával is rímpárt alkot, ugyanakkor a program jelenlegi elemzésében nem alkot rímpárt a negyedik sor sorvégi szavával, vagyis egy rímelő szó hívórímként mindig csak a hozzá legközelebbi rímelő szóval alkothat rímpárt. Ezzel a fajta elemzéssel az vizsgálható, hogy egy adott típusú rímszó milyen típusú, a rímszerkezetben azt közvetlenül követő (vagy megelőző) rímszót hív elő a legvalószínűbben. Ennek például a rímelés pszicholingvisztikai irányultságú vizsgálataiban lehet jelentősége. A rímek automatikus elemzését a későbbiekben ugyanakkor érdemes lenne úgy továbbfejleszteni, hogy a bokorrímek és az egyéb, kettőnél több elemű rímek esetében ne csak az egymást követő tagok által alkotott rímpárokat lehessen vizsgálni, illetve hogy a kettőnél több tagú szerkezeteket egy egységként is le lehessen kérdezni. Ez utóbbi megoldás segíthetné a rímelés összetettebb szerkezeteinek a leírására irányuló vizsgálatokat. A rímek automatikus annotálása egy megfelelő méretű korpuszban – a tervek szerint az ELTE Verskorpusz ilyen lenne – lehetőséget adhat egy rímszótár elkészítésére is.<sup>19</sup>

A *hunpoem\_analyzer-TEI* program úgyszintén elemzi a sorok ritmusát. Szemben az angol és egyéb indoeurópai nyelvű versek szóhangsúlyra épülő versritmusával, a magyar időmértékes versritmus a rövid és hosszú szótagok különbségére épül. A sorok hosszú és rövid szótagjainak az annotálása néhány egyszerű, a magyar verstanban jól ismert szabály alapján elvégezhető, így az előző részben bemutatott példáktól eltérően nem szükséges kiejtésszótárakat beépíteni az algoritmusba. Ezek a szabályok a következők: (1) a program rövid szótagként elemzi azokat a szótagokat, amelyekben rövid magánhangzó van, és közvetlenül a rövid magánhangzó után nem áll mássalhangzó, vagy csak egy rövid mássalhangzó áll; (2) a program hosszú szótagként elemzi

<sup>18</sup> Az asszonáncoknak ilyen fokozati leírását adja: Arany János, „Valami az asszonáncról,” in *Arany János összes művei X. kötet. Prózai művek 1.*, szerk. Keresztury Dezső (Budapest: Akadémiai Kiadó, 1962), 213–217. Arany János leírásának megkülönböztető jegyekre épülő, fonológiai keretben történő újraértelmezését adja: Szépe György, „Nyelvészeti jegyzetek Arany Jánosnak »Valami az asszonáncról« című tanulmányáról,” *Magyar Nyelvőr* 93, 1. sz. (1969): 1–32. Funkcionális kognitív nyelvészeti keretben a magyar rím úgyszintén Arany leírásából kiinduló, prototípusalapú szerveződését vázolja fel: Simon Gábor, „A magyar rím fonológiai leírása funkcionális-kognitív megközelítésben,” *Magyar Nyelvőr* 136, 2. sz. (2012): 65–82, valamint Simon Gábor, *Egy kognitív poétikai rímelmélet megalapozása* (Budapest: Tinta Könyvkiadó, 2014).

<sup>19</sup> Számítógépes nyelvészeti eljárásokat használó rímszótár elkészítésének módszeréről lásd Mártonfi Attila, „Egy magyar rímszótár terve,” in „*Mielz valt mesure / que ne fait estultie*”: *A hatvanéves Horváth Iván tiszteletére*, szerk., Bartók István et al. (Budapest: Krónika Nova Kiadó, 2008), 198–204.

azokat a szótagokat, amelyekben hosszú magánhangzó áll, valamint azokat a rövid magánhangzós szótagokat, amelyekben hosszú vagy egynél több mássalhangzó követi a magánhangzót. Az elemzés kimenete minden sor esetében egy 0 és 1 karakterekből álló karaktersor, amelyben a 0 a rövid, az 1 pedig a hosszú szótagokat reprezentálja (pl. „*Nincsen apám, se anyám*” – 1001001). Ezt az elemzést a program automatikusan elvégzi minden versen, függetlenül attól, hogy jellemző-e rá valamilyen szabályos időmértékes versritmus vagy nem. Az így megadott ritmusból a program jelen állapotában nem kísérel meg valamilyen, a versre általánosan jellemző metrumot megállapítani, és verslábakat sem határoz meg. A távlati tervek között szerepel, hogy az előző részben bemutatott példákhoz hasonlóan a program az időmértékes ritmusból absztrahálni tudjon metrumot, illetve hogy az időmértékes verselés mellett az ütemhangsúlyos (és a szimultán) verselést is felismerje.<sup>20</sup>

A *hunpoem\_analyzer-TEI* program utolsó funkciója az alliterációk automatikus felismerése. Az alliteráció tágabb értelmezése alapján nem csupán az azonos mássalhangzóval, hanem az azonos magánhangzóval kezdődő szavak is alliterációként annotálódnak. A program jelenlegi változata csak versszakon belül elemez alliterációkat. Ez tehát azt jelenti, hogy egy versszak utolsó és a következő versszak első, azonos hanggal kezdődő szava nem annotálódik alliterációként. A program nem csupán azokat a szerkezeteket elemzi alliterációként, amelyekben egymást követő szavak ugyanazzal a hanggal kezdődnek, hanem azokat is, amelyekben két ugyanolyan hanggal kezdődő szó közé beékelődik egy másik hanggal kezdődő szó. Így például Babitsnak a „*Bus donna barna balkonon*” sorát egy egységként elemezné mint alliterációt. Minden alliterációként elemzett szerkezet kap egy *a* és *n* betűből álló karaktersort, amelyben az *a* betű az egymással alliteráló szavakat, az *n* betű pedig az alliteráló szavak közé beékelődő nem alliteráló szavakat jelöli (pl. „*Bus donna barna balkonon*” - anaa). A programba egyelőre nincs beépítve *stopword*-lista, aminek következtében az *az asztal* típusú szerkezetek is alliterációként annotálódnak.

A *hunpoem\_analyzer-TEI* programnak nemcsak a bemenete, hanem az elemzett tulajdonságokat tartalmazó kimenete is megfelel a TEI XML-szabványnak.

#### 4. A József Attila-korpusz létrehozásának és automatikus elemzésének főbb lépései

A József Attila verseit tartalmazó korpuszt a Magyar Elektronikus Könyvtár (MEK)<sup>21</sup> oldaláról letölthető *József Attila összes költeménye* című, MEK-00708 azonosítószámú digitális dokumentumból hoztam létre.<sup>22</sup> A töredékeket, az idegen nyelvű verseket,

<sup>20</sup> Az időmértékes metrum felismerése első megközelítésben a jambikus, trochaikus, anapestikus és daktilikus általános kategóriákba történő automatikus besorolást jelenthetné. Az ütemhangsúlyos verselés automatikus felismerésének előfeltétele a szóhangsúlyok automatikus annotálása. A versritmusról például lásd: Szepes Erika és Szerdahelyi István, *Verstan* (Budapest: Gondolat Kiadó, 1981).

<sup>21</sup> Magyar Elektronikus Könyvtár, hozzáférés: 2019.01.19, <https://mek.oszk.hu/>.

<sup>22</sup> Az ELTE Verskorpusz építése során alapvetően a MEK adatbázisában található versgyűjteményeket használjuk, amelyek többnyire egységes formátumban vannak, így nem kell minden egyes szerző esetében külön szkriptet írni a versek szerkezeti egységeinek az annotációit tartalmazó TEI XML-fájlok létrehozására, amelyek a további annotálási lépések bemenetét adják. Bár a <http://jozsef.atta.elte.hu/> oldalon jobb minőségben szerepelnek a szövegek, a MEK-szövegkiadás használ-

valamint a *Változatok* és a *Rögtönzések* címszó alatt szereplő szövegeket mellőzve 532 vers került a korpuszba. Egy XQuery szkripttel minden versből létrehoztam egy, a versek szerkezeti egységeinek (cím, versszakok, sorok) az *inline* annotációit tartalmazó TEI XML-fájlt. Ezt követően a versek versszakbeosztását ellenőriztem a kritikai kiadás alapján, és ahol eltérés volt, ott javítottam.<sup>23</sup> Ez a rímképlet és a rímpárok pontosabb megállapítása miatt volt fontos. Egyéb szempontból nem javítottam a szövegeket. Következő lépésben a TEI XML-fájlokon lefuttattam a *hunpoem\_analyzer-TEI* programot, illetve az ebbe beépített *e-magyar* program *emtsv* változatát. Ennek eredményeként a TEI XML-fájlokba bekerültek a szavak grammatikai tulajdonságai, azaz a szavak lemmája, szófaja és morfoszintaktikai jellemzői, valamint az előző pontban részletezett, vershangzáshoz kapcsolódó annotációk, vagyis a szavak főbb fonológiai jellemzői, a versszakok rímképlete, a rímpárok, a sorok ritmusa és az alliterációk elemzése. Ezt követően a TEI XML-fájlok formátumán egy XSLT stíluslappal kisebb módosításokat hajtottam végre, mivel a TEI XML-formátum alapvetően szövegek szabványos tárolására és megosztására lett kitalálva, nem pedig arra, hogy bonyolultabb lekérdezéseket is egyszerűbben meg lehessen írni, illetve gyorsan le lehessen futtatni. A kimenetül kapott XML-fájlokat az *eXist-db* nevű XML-adatbázis-kezelő programba töltöttem be, ahol XQuery nyelven meg lehet írni a lekérdezéseket. Hangsúlyozandó tehát, hogy az alábbiakban szereplő gyakorisági listákat nem az *e-magyar*, és nem is a *hunpoem\_analyzer-TEI* program hozza létre, hiszen ezek funkciója pusztán egy szó vagy egy több szóból álló szerkezeti egység (sor, versszak stb.) elemzése, illetve az elemzésnek az XML-fájlba való beírása. A különböző gyakorisági listák létrehozása az XQuery-lekérdezésekbe lett beépítve.

A József Attila-korpusz alapját adó MEK-dokumentum a címe ellenére természetesen nem tartalmazza az összes József Attila-verset, és a szövegek is több esetben eltérnek a kritikai kiadás alapszövegétől. Ennek a tanulmánynak az elsődleges célja azonban nem az, hogy minél pontosabb kvantitatív adatokat szolgáltatson József Attila költészetéről, hanem az, hogy József Attila verseinek a példáján keresztül bemutassa a vershangzáshoz kapcsolódó jellemzők automatikus feltárásának egy módszerét.<sup>24</sup> Mindazonáltal feltételezhető (vagy legalábbis remélhető), hogy a korpuszban nem szereplő, illetve rontott szöveggel szereplő versek nem befolyásolják lényegesen az alábbiakban bemutatandó gyakorisági listák elemeinek a sorrendjét.

## 5. A József Attila-versek szókészletének néhány jellemzője

A versek automatikusan feltárt hangzásjellemzőinek az ismertetése előtt érdemes a korpusz főbb lexikai tulajdonságaiból is bemutatni párat. A lexikai tulajdonságok

latával részben az volt a célom, hogy teszteljem a készülő ELTE Verskorpusz számára kialakított, a MEK adatbázisában található szövegekre épülő annotációs folyamatot.

<sup>23</sup> József Attila, *József Attila összes versei*, kiad. Stoll Béla (Budapest: Balassi Kiadó, 2005).

<sup>24</sup> A József Attila költészetéről pontosabb adatokat szolgáltató, a különböző szövegváltozatokat is figyelembe vevő írói szótár tervezetéről lásd: Mártonfi Attila, „Számítógép és írói szótár – különös tekintettel a készülő József Attila szótárra,” *Magyar Nyelv* 110, 1. sz. (2014): 30–46.

automatikus kinyerését az *e-magyar*nak a verseken való lefuttatása tette lehetővé. Az alábbi 1. táblázat a korpusz legfontosabb mennyiségi jellemzőit mutatja be.<sup>25</sup>

versek száma	532
szavak száma	58144
lemmák száma	10038

1. táblázat. A korpusz legfontosabb mennyiségi jellemzői

Egy szó lemmája a szó szótári alakja. Például a *futott*, *futnánk*, *fut*, *fuss* szavak lemmája a *fut* szó. A lemmák száma tehát a korpusz szókincsének a méretét mutatja meg. A 2. táblázat a korpuszban szereplő húsz leggyakoribb lemmát mutatja be. A lemmák utáni oszlopban adtam meg az előfordulási számot, a később szereplő gyakorisági listáknál is ezt a módszert követem. A táblázatban egy adott lemmához különböző szófajú, adott esetben homonim szóalakok is tartozhatnak, például a *ki* lemma a névmási és az igekötői szóalakokat is magában foglalja.

1	a	4174
2	az	1432
3	s	1260
4	és	877
5	nem	844
6	én	768
7	van	742
8	is	585
9	hogy	553
10	mint	525
11	ha	483
12	csak	384
13	de	375
14	ki	325
15	egy	323
16	már	319
17	meg	303
18	te	298
19	ő	291
20	mi	289

2. táblázat. A leggyakoribb lemmák

Látható, hogy a húsz leggyakoribb lemma között többnyire névelők, kötőszavak, névmások szerepelnek, azaz olyan fogalmilag kevésbé tartalmas szavak, amelyek nem túl sokat árulnak el a szöveg jellegéről. Érdekes tehát fogalmilag tartalmasabb jelentésű

<sup>25</sup> A verscímeken nem lett lefuttatva az *e-magyar* és a vershangzást elemző program, így ezek nem alkotják részét a tanulmányban bemutatott adatoknak.

szófajokhoz kapcsolódó szólistákat lekérdezni. A 3. táblázat a korpuszban szereplő tíz leggyakoribb igei, főnévi és melléknévi lemmát mutatja be.

**Ige**

1	van	742
2	lesz	273
3	tud	192
4	szeret	172
5	lát	143
6	nincs	130
7	jön	130
8	kell	127
9	vár	113
10	él	112

**Főnév**

1	szem	179
2	szív	172
3	ég	170
4	föld	162
5	ember	160
6	lélek	147
7	világ	129
8	isten	111
9	szó	102
10	kéz	98

**Melléknév**

1	szép	211
2	nagy	208
3	jó	155
4	kis	97
5	bús	74
6	erős	65
7	tiszta	58
8	friss	52
9	lágý	49
10	lassú	48

3. táblázat. A leggyakoribb igei, főnévi és melléknévi lemmák.

Bár ennek a tanulmánynak nem célja a különböző lekérdezett listák elemzése, talán érdemes megjegyezni azt a számomra meglepő eredményt, hogy a tíz leggyakoribb melléknév többsége pozitív konnotációjú. Ez persze összefügghet azzal a nyelvi univerzáléval, hogy a minőségre vonatkozó kategóriák megnevezésére jellemzően pozitív konnotációjú mellékneveket használunk.



Természetesen specifikusabb nyelvtani kategóriák szólistáit is le lehet kérdezni, az alábbi, 4. táblázatban például a tíz leggyakoribb E/1, E/2 és E/3 alakban előforduló igei lemma szerepel. Félkövérral emeltem ki azokat az igeiket, amelyek a háromból csak az egyik kategóriában szerepelnek a leggyakoribb tíz igei lemma között.

	E/1		E/2		E/3	
1	van	186	van	48	van	401
2	tud	96	lesz	44	lesz	134
3	szeret	70	jön	39	nincs	128
4	lesz	52	néz	28	kell	122
5	lát	50	szeret	27	száll	66
6	vár	29	mond	25	jön	56
7	hisz	28	tud	23	hull	47
8	érez	26	ad	22	vár	45
9	megy	24	lát	22	volna	45
10	él	24	alszik	22	tud	44

4. táblázat. Az E/1, E/2 és E/3 alakban leggyakrabban előforduló igei lemmák.

Nemcsak szólisták, hanem a szavaknál absztraktabb, grammatikai kategóriák is lekérdezhetők. Az 5. táblázat a korpusz leggyakoribb tíz szófaját mutatja be.<sup>26</sup>

1	főnév	15774
2	ige	11613
3	melléknév	6984
4	határozószó	6184
5	determináns	5588
6	névmás	4896
7	mellérendelő kötőszó	3131
8	alárendelő kötőszó	1716
9	névutó	685
10	tulajdonnév	523

5. táblázat. A leggyakoribb szófajok.

## 6. A József Attila-versek automatikusan feltárt hangzásjellemzői

Az előző részben bemutatott, szavakat és grammatikai kategóriákat tartalmazó gyakorisági listák minden bizonnyal nem jelentenek nagy újdonságot azok számára,

<sup>26</sup> Az *e-magyar*nak az UD (Universal Dependencies) típusú szófaji és morfoszintaktikai címkéket kiadó kimenetét használtam. Az UD-szabványt követő *e-magyar* nem minden esetben úgy osztja ki a szófaji címkéket, ahogyan azt a magyar leíró nyelvtanok teszik. Az *e-magyar* UD kimenetében szereplő eredeti szófaji címkék és azoknak az általam használt fordításai a következők: NOUN – főnév, VERB – ige, ADJ – melléknév, ADV – határozószó, DET – determináns, PRON – névmás, CONJ – mellérendelő kötőszó, SCONJ – alárendelő kötőszó, ADP – névutó, PROP – tulajdonnév, NUM – számnév. Az UD szabvány magyar nyelvre való alkalmazását lásd az alábbi címen, hozzáférés: 2020.01.19, [https://universaldependencies.org/treebanks/hu\\_szeged/index.html](https://universaldependencies.org/treebanks/hu_szeged/index.html).

akik foglalkoznak korpusznyelvészettel, hiszen ilyen gyakorisági listákat a nyelvileg elemzett korpuszok lekérdezőfelületei általában tudnak generálni (erre jó példa a már említett Magyar Nemzeti Szövegtár lekérdezőfelülete<sup>27</sup>). Egy szövegcsoporthoz kvantitatív módon azonban nem csak a szókészletre vonatkozó tulajdonságok mentén jellemezhetünk. Egy lírai szövegeket tartalmazó korpuszt például az alapján is leírhatunk, hogy melyek a legtipikusabb, leggyakoribb rímképletek, rímpárok, ritmusok, vagy éppen alliterációtípusok. A továbbiakban olyan gyakorisági listákat mutatok be, amelyek lekérdezését a vershangzáshoz kapcsolódó tulajdonságok előzetes, automatikus annotálása tette lehetővé.

Egy verskorpusz kapcsán feltehetjük azt az egyszerű kérdést, hogy melyek a leggyakoribb versszakszámok. A 6. táblázat a tíz leggyakoribb versszakszámot mutatja be.

Versszakszám

1	4	152
2	3	82
3	1	66
4	5	54
5	6	46
6	2	45
7	7	21
8	8	17
9	9	14
10	10	8

6. táblázat. A leggyakoribb versszakszámok.

A táblázatból láthatjuk, hogy a négy versszakból álló versek a leggyakoribbak, ezekből 152 darab szerepel a korpuszban. A lista második helyén pedig a három versszakos versek állnak.

Egy verskorpusz egyik fontos jellemzője az is, hogy a versszakok esetében melyek a legtipikusabb sorszámok és rímképletek. A 7. táblázat a versszakokra jellemző leggyakoribb tíz sorszámot és rímképletet mutatja be.

Sorszám

1	4	1301
2	3	348
3	2	251
4	5	142
5	8	139
6	6	88
7	1	69
8	7	64
9	10	32

<sup>27</sup> Magyar Nemzeti Szövegtár, hozzáférés: 2020.01.19, <http://clara.nytud.hu/mnsz2-dev/>.

10	11	16
11	9	16
12	12	16

Rímképlet

1	abab	445
2	aabb	242
3	abcb	158
4	abba	143
5	aab	142
6	aa	138
7	ab	113
8	aba	111
9	aaaa	82
10	#	69

7. táblázat. A versszakok leggyakoribb sorszámai és rímképletei.

A táblázatból látható, hogy a négysoros versszakok a legtipikusabbak, ezekből 1301 darab szerepel a korpuszban, a második leggyakoribb típus pedig a háromsoros versszak. A leggyakoribb rímképlet az abab típus, azaz a négysoros keresztrímek, amelyből 445 szerepel a korpuszban. A lista második helyén pedig az aabb típus áll, vagyis a négysoros párosrímek. A rímképletek oszlopban a 10. helyen szereplő # jel az egysoros versszakokra utal.

Nemcsak versszakok, hanem verssorok kapcsán is lekérdezhetünk különböző gyakorisági listákat. A 8. táblázat a korpusz soraira jellemző tíz leggyakoribb szószámot, szótagszámot és ritmusképletet mutatja be. A ritmusképletekben a 0 a rövid szótag, az 1 pedig a hosszú szótag jele.<sup>28</sup>

Szószám

1	5	3072
2	4	2965
3	6	1891
4	3	1672
5	7	914
6	2	574
7	8	365
8	9	164
9	1	136
10	10	72

<sup>28</sup> A sorok utolsó szótagjának az elemzésében a program nem veszi figyelembe a következő sor elejét.

Szótagszám

1	10	2509
2	8	2296
3	9	2229
4	11	1852
5	7	708
6	6	509
7	12	383
8	4	243
9	13	205
10	5	166

Ritmus

1	110101010	57
2	011101010	55
3	010111010	48
4	010101010	46
5	11010101010	46
6	11011100	46
7	111101010	44
8	01111100	43
9	110111010	42
10	01011101	41
11	01111101	41

8. táblázat. A sorok leggyakoribb szószámai, szótagszámai és ritmusképletei.

A vizsgált korpuszban a leggyakoribb tehát az öt szóból álló sor, szótagszám tekintetében pedig a tíz szótagos sor. A leggyakoribb ritmusképlet viszont kilenc szótagból áll, hagyományos jelöléssel: – – U – U – U – U. Megjegyzendő, hogy az első négy ritmusképlet valójában mind ötödfeles jambus. A különbség pusztán annyi, hogy a jambusokat helyettesítő spondeusok máshol jelennek meg. A többi ritmusképlet is mind jambikus lejtésű. Ahogyan arra már utaltam, az annotáló program továbbfejlesztésének a távlati céljai között szerepel a ritmusból metrumra történő absztrahálás funkciójának a beépítése. Más szerzők verseivel való összehasonlításban érdekes lehet a hosszú és rövid szótagok korpuszbeli előfordulásának a száma is, ezt az alábbi, 9. táblázat mutatja be.

hosszú szótagok száma	62731
rövid szótagok száma	51908

9. táblázat. A hosszú és rövid szótagok száma.

Ahogyan arról a tanulmány elején szó volt, a versek automatikus elemzéséhez fejlesztett *hunpoem\_analyzer-TEI* program a szavak szótagszámát, hangrendjét és fonológiai szerkezetét is elemzi. A 10. táblázat a magas, mély és vegyes hangrendű

szavak gyakorisági sorrendjét, valamint a szavak szótagszámának a gyakorisági sorrendjét mutatja be. A szótagszám oszlopban a 0 érték ötödik helyen való szereplése a s kötőszó gyakori használatával magyarázható.

Hangrend

1	magas	25452
2	mély	23908
3	vegyes	7461

Szótagszám

1	1	21176
2	2	19608
3	3	10985
4	4	4134
5	0	1272
6	5	773
7	6	125
8	7	19
9	8	1

10. táblázat. A szavak hangrendjének és szótagszámainak a gyakorisági listája.

A 11. táblázat a tíz leggyakoribb fonológiai szerkezetet mutatja be. A fonológiai szerkezetek reprezentációiban a C a mássalhangzókat, a V pedig a magánhangzókat jelöli, a V után álló F az elöl képzett, a B pedig a hátul képzett magánhangzókra utal. Az ezt követő 1 szám arra utal, hogy a magánhangzó rövid, a 2 pedig, hogy a magánhangzó hosszú. A VB1 szerkezetnek az első helyen szereplése tehát az *a* névelőnek köszönhető.

Fonológiai szerkezet

1	VB1	4174
2	C, VB1, C	2000
3	C, VF1, C	1911
4	VF2, C	1668
5	C, VF1, C, C	1634
6	VB1, C	1393
7	C, VF1	1362
8	C, VF1, C, VF1, C	1348
9	C, VB1, C, VB1, C	1300
10	VF1, C	1290

11. táblázat. A szavak leggyakoribb fonológiai szerkezetei.

A korpuszban ugyancsak annotálva lettek az alliterációk, így ezekre vonatkozó gyakorisági listákat is le lehetett kérdezni. A 12. táblázat a tíz leggyakoribb alliterációként elemzett szerkezetet mutatja be. Az *a* betű jelöli a szerkezetben egymással alliteráló szavakat, az *n* pedig az alliteráló szavak közé esetlegesen beékelődő nem

alliteráló szavakat. A *hunpoem\_analyzer-TEI* program csak azokat a szerkezeteket elemzi alliterációként, amelyekben két alliteráló szó közé maximum egy nem alliteráló szó ékelődik be. A programba egyelőre nincs beépítve stopword-lista, így az *az asztal* típusú szerkezetek is részei a listának.

Szerkezet

1	ana	2237
2	aa	2155
3	anana	157
4	aaa	147
5	aana	133
6	anaa	120
7	ananana	16
8	anaaa	10
9	ananaa	10
10	aaaa	10

12. táblázat. Az alliterációk leggyakoribb típusai.

A 13. táblázat az alliterációk leggyakoribb típusait mutatja be szófajok alapján. Az első oszlopban a beékelődő, nem alliteráló szavakat nem tartalmazó szerkezetek leggyakoribb tíz típusa szerepel az alliterációt alkotó szavak szófaja alapján. Ebből a listából – kompenzálva a *stopword*ök hiányát – kihagytam azokat a kételemű alliterációtípusokat, amelyek determinánsokat tartalmaznak. A másik oszlop a leggyakoribb tíz, beékelődő szót tartalmazó szerkezetet mutatja be a szavak szófaja alapján. Az alliteráló szavak közé beékelődő szó szófaját félkövérrel emeltem ki. Ebből a listából is kihagytam azokat a – beékelődő szóval együtt – háromelemű alliterációtípusokat, amelyekben a determináns nem beékelődő, hanem alliteráló szóként jelenik meg.

Szófaj – nincs benne nem alliteráló szó

1	melléknév, főnév	193
2	főnév, főnév	170
3	főnév, ige	166
4	ige, ige	97
5	ige, névmás	70
6	ige, főnév	64
7	határozószó, ige	63
8	ige, határozószó	58
9	főnév, melléknév	56
10	melléknév, melléknév	55

Szófaj – beékelődő nem alliteráló szóval

1	főnév, <b>determináns</b> , főnév	76
2	ige, <b>determináns</b> , főnév	74
3	melléknév, főnév, ige	49

4	főnév, melléknév, főnév	43
5	főnév, főnév, ige	34
6	határozószó, ige, határozószó	34
7	ige, határozószó, ige	32
8	főnév, főnév, főnév	30
9	ige, melléknév, főnév	29
10	melléknév, főnév, főnév	29

13. táblázat. Az alliterációk leggyakoribb típusai szófajok alapján.

A táblázatból látható, hogy a beékelődő szavakat nem tartalmazó leggyakoribb tíz szerkezet között nincsen olyan szófaj-kombináció, amely kettőnél több elemű lenne, és a beékelődő szavakat tartalmazó leggyakoribb tíz szerkezet között sincs olyan, amelyben kettőnél több alliteráló szó szerepelne.

A következő típusa az itt bemutatandó, vershangzásra vonatkozó kvantitatív jellemzőknek a rímhelyzetben lévő szavakhoz kapcsolódó gyakorisági listák. A rímnek nem csupán fonológiai, hanem szemantikai funkciója is van. A rímhelyzet a versvilág felépítése szempontjából jelentéstanilag kitüntetett pozíció, így egy verskorpusz esetében érdekes lehet, hogy milyen típusú szavak kerülnek leggyakrabban rímhelyzetbe.<sup>29</sup> A 14. táblázat a József Attila-korpuszban szereplő tíz leggyakoribb rímhelyzetben lévő tokent (szóalak), lemmát (szótári alak) és szófajt mutatja be. Ezen listák esetében tehát nincsen annak jelentősége, hogy a szó hívó- vagy felelőrim pozíciót tölt be, vagy esetleg mindkettőt.

#### Token

1	is	53
2	el	44
3	meg	39
4	már	35
5	van	34
6	vagyok	32
7	én	30
8	alatt	25
9	engem	25
10	ég	22

#### Lemma

1	van	168
2	én	133
3	ég	74
4	maga	61

<sup>29</sup> A rím jelentéstani szerepét mutatja be Simon, *Egy kognitív poétikai rímelmélet megalapozása*. A monográfiában szerepel egy József Attila-verseket is tartalmazó verskorpusz rímhelyzetben lévő szavainak a manuálisan létrehozott szófaji annotációira épülő vizsgálat.

5	lesz	57
6	világ	56
7	is	53
8	te	50
9	szeret	50
10	ő	46

**Szófaj**

1	főnév	5002
2	ige	3018
3	melléknév	810
4	határozószó	728
5	névmás	510
6	névutó	223
7	tulajdonnév	77
8	számnév	44
9	mellérendelő kötőszó	37
10	PART <sup>30</sup>	37

**14. táblázat.** A leggyakoribb rímhelyzetben lévő tokenek, lemmák és szófajok.

Érdemes megjegyezni, hogy míg a korpusz leggyakoribb húsz lemmája között egyáltalán nem szerepelt főnév, és igeiként is csak a *van* lemma szerepelt, addig a leggyakoribb tíz rímhelyzetben lévő lemma között a *van* mellett szerepel a *lesz* és a *szeret* ige, valamint a *világ* főnév és az *ég* lemma, amely főnév és ige is lehet. Ha a szófajok gyakorisági listáját nézzük, akkor megállapíthatjuk, hogy a rímhelyzetben lévő első négy leggyakoribb szófaj (főnév, ige, melléknév, határozószó) ugyanaz, mint a korpusz összes szava esetében. Ezt követően változik a sorrend, például nincs a leggyakoribb tíz rímhelyzetben lévő szófaj között determináns és alárendelő kötőszó, a mellérendelő kötőszó pedig hátrébb került. Ezek a változások persze nem meglepőek, hiszen a versek sorai a legtöbb esetben tagmondatok, amelyek végén nincsen névelő vagy kötőszó.

A rímhelyzetben lévő szavak esetében is lekérdezhettük azok fonológiai tulajdonságait. A 15. táblázat a rímhelyzetben lévő szavak hangrendjének és szótagszámának a gyakorisági listáját mutatja be.

**Hangrend**

1	magas	4804
2	mély	3939
3	vegyes	1749

<sup>30</sup> Az *e-magyar* UD típusú kimenetében a PART (particle) címkét kizárólag a *meg* igeikötő kapja. Ez a címke tehát nem egyezik meg a magyar leíró nyelvészetben használt partikula szófajának a kategóriájával. A többi igeikötő az ADV (határozószó) címkét kapja.



Szótagszám

1	2	4439
2	3	3076
3	4	1497
4	1	1140
5	5	292
6	6	44
7	7	4

15. táblázat. A rímhelyzetben lévő szavak hangrendjének és szótagszámának gyakorisági listája.

A 15. táblázatból kiderül, hogy a mély hangrendű rímhelyzetben lévő szavak aránya csökken, míg a vegyes hangrendű szavak aránya nő az összes szó mély és vegyes hangrendű szavának az arányához képest. Az összes szónak 42,1%-a mély hangrendű és 13,1%-a vegyes hangrendű. Ezzel szemben a rímhelyzetben lévő szavaknak 37,5%-a mély hangrendű és 16,7%-a vegyes hangrendű. A jövőben érdemes lenne megvizsgálni, hogy a hangrendi arányoknak ez a változása a magyar rímelésre jellemző általános, azaz más szerzőknél is megjelenő mintázatnak tekinthető-e. A rímhelyzetben lévő szavak szótagszámának a gyakorisági listája egy dologban különbözik a korpusz összes szavára vonatkozó gyakorisági listától: az egy szótagos szavak hátrébb kerültek a listában. Ez minden bizonnyal azzal a szófajok kapcsán megállapított jellemzővel magyarázható, hogy az egy szótagú névelők nem állnak a verssorok végén, és a gyakran egy szótagú kötőszavak sem jellemzőek ebben a pozícióban.

A 16. táblázat a rímhelyzetben lévő szavak leggyakoribb tíz fonológiai szerkezetét mutatja be.

Fonológiai szerkezet

1	C, VB1, C, VB1, C	394
2	C, VF1, C, VF1, C	350
3	C, VF1, C, C, VF1, C	324
4	C, VB1, C, C, VB1, C	189
5	C, VB1, C, C, VB1	160
6	C, VB2, C	158
7	C, VF1, C, VF1, C, VF1, C	153
8	C, VF1, C, VF1, C, C, VF1, C	152
9	C, VF1, C, VF2, C	145
10	C, VB1, C, VB2, C	142

16. táblázat. A rímhelyzetben lévő szavak leggyakoribb fonológiai szerkezetei.

A szótagszámnak a fentebb említett változása magyarázza a leggyakoribb tíz fonológiai szerkezet listájában tapasztalható jelentősebb eltéréseket a korpusz összes szavára vonatkozó listához képest. A rímhelyzetben lévő szavak esetében az első öt leggyakoribb fonológiai szerkezet két magánhangzót tartalmaz, és a leggyakoribb tíz szerkezet között is csak egy darab egy magánhangzót tartalmazó szerkezet szerepel,

ugyanakkor két darab három magánhangzót tartalmazó szerkezet is megjelenik. Ezzel szemben az összes szóra vonatkozó listában csupán két darab két magánhangzót tartalmazó szerkezet szerepel a lista nyolcadik és kilencedik helyén, az összes többi szerkezet egy szótagú.

Végezetül érdekes lehet az is, hogy melyek a leggyakoribb rímpárok, illetve rímpártípusok. A 17. táblázat első oszlopa a korpuszban legalább háromszor szereplő rímpárokat mutatja be gyakoriságuk sorrendjében, a második oszlop pedig a rímpárok lemmatizált alakjaiból tünteti fel azokat, amelyekből legalább három szerepel a korpuszban. Hangsúlyozandó, hogy a táblázatban szereplő gyakorisági listák figyelembe veszik a rímpárt alkotó szavak sorrendjét. Például a *tenger-ember* rímpárba nincsenek beleszámolva az *ember-tenger* sorrendű rímpárok, ezek a lista külön tételét alkotják.

Token

1	tenger – ember	4
2	szemegödre – mindörökre	4
3	vagyok – csillagok	3
4	az – igaz	3
5	szellem – ellen	3
6	hamis – is	3
7	pajtás – hajtás	3
8	magam – van	3
9	lelkem – telken	3
10	össze – fürösze	3
11	szerelem – velem	3
12	kell – el	3
13	apámat – Amerikának	3
14	agyunk – vagyunk	3

Lemma

1	maga – van	5
2	lélek – telek	5
3	tenger – ember	5
4	szemegöd – mindörökre	4
5	van – csillag	3
6	hisz – visz	3
7	között – köd	3
8	az – igaz	3
9	tör – meggyötör	3
10	szellem – ellen	3
11	elme – szerelem	3
12	maga – szó	3
13	vágy – ágy	3
14	hamis – is	3
15	pajtás – hajtás	3
16	össze – füröszt	3

17	szerelem – én	3
18	kell – el	3
19	test – este	3
20	apa – Amerika	3
21	van – maga	3
22	agy – van	3

17. táblázat. Rímpárok gyakorisági listája tokenek és lemmák alapján – a rímpárok tagjainak a sorrendje számít.

A táblázat kapcsán megjegyzendő, hogy sok esetben bizonyos rímpároknak a többszöri előfordulása kizárólag egy verssel magyarázható. Például a *szemegödre–mindörökre* rímpárnak mind a négy előfordulása a *Bús magyar éneke* című versben található, az *apámat-Amerikának* rímpárnak pedig mind a három előfordulása a „Csak most...” kezdetű versben jelenik meg.

A 18. táblázatban a leggyakoribb tíz, rímpárt alkotó szófaj-kombináció, valamint a leggyakoribb tizenöt, rímpárt alkotó szótagszám-kombináció szerepel. Ezekben az esetekben is számít a rímpárok tagjainak a sorrendje.

Szófaj

1	főnév – főnév	1312
2	főnév – ige	624
3	ige – ige	608
4	ige – főnév	601
5	melléknév – főnév	203
6	határozószó – főnév	175
7	főnév – melléknév	169
8	főnév – határozószó	156
9	névmás – főnév	123
10	főnév – névmás	113

Szótagszám

1	2 – 2	1088
2	2 – 3	624
3	3 – 3	590
4	3 – 2	520
5	2 – 4	335
6	4 – 2	266
7	3 – 4	211
8	1 – 2	201
9	4 – 4	192
10	1 – 1	177
11	2 – 1	176
12	1 – 3	173

13	4 – 3	153
14	3 – 1	119
15	2 – 5	77

18. táblázat. Rímpárok gyakorisági listája szófajok és szótagszámok alapján – a rímpárok tagjainak a sorrendje számít.

A 18. táblázatban szereplő szótagszámok kapcsán egy érdekességre érdemes felhívni a figyelmet. Egy különböző szótagszámokat tartalmazó kombináció esetében mindig több előfordulással jelenik meg az a változat, amelyben az első tag, vagyis a hívórím a kevesebb szótagszámú. Vagyis a gyakorisági listában a 2-3 kombináció előrébb szerepel, mint a 3-2 kombináció, a 2-4 kombináció előrébb szerepel, mint a 4-2 kombináció, a 3-4, az 1-2, az 1-3 és a 2-5 pedig előrébb szerepel, mint a 4-3, a 2-1, a 3-1 és az 5-2 kombinációk. Ebből persze még nem következik, hogy egy József Attilára specifikusan jellemző mintázatot találtunk, hiszen könnyen lehet, hogy egy, a rímhez kapcsolódó általános, más szerzőkre is jellemző sémáról van szó, amely valamilyen kognitív, pszicholingvisztikai tényezővel magyarázható. A kérdés eldöntéséhez a jövőben érdemes lenne más szerzők verseit is megvizsgálni ebből a szempontból.

Az előző, 17. és 18. táblázatok gyakorisági listái figyelembe vették a rímpárok elemeinek a sorrendjét. Bizonyos kutatási kérdések megválaszolásához azonban szükséges lehet olyan listák előállítás is, amelyek nem veszik figyelembe a sorrendet. Ezt mutatja be a 19. és a 20. táblázat. A 19. táblázat két oszlopában a korpuszban legalább négyszer előforduló rímpárokat, illetve lemmatizált rímpárokat tüntettem fel, a rímpárok tagjainak a sorrendjét figyelmen kívül hagyva (vagyis az *ember-tenger* vagy a *maga-van* rímpár a *tenger-ember* és a *van-maga* kombinációkat is magában foglalja).

Token

1	ember – tenger	6
2	hamis – is	5
3	magam – van	5
4	engem – szívemben <sup>31</sup>	4
5	is – mégis	4
6	lelkem – telken	4
7	végtelenbe – egyre	4
8	agyunk – vagyunk	4
9	szerelem – velem	4
10	költemény – én	4
11	mindörökre – szemegödre	4

<sup>31</sup> A különböző írásváltozatú alakok természetesen különböző tokenekként kerülnek be a listába. Ugyanakkor az *engem-szívemben* rímpár esetében rákerestem az *engem-szívemben* rímpárra, amelyből egyet találtam, ezzel együtt tehát összesen öt előfordulásról beszélhetünk.

Lemma

1	maga – van	8
2	ember – tenger	7
3	lélek – telek	6
4	maga – szó	5
5	agy – van	5
6	hamis – is	5
7	hisz – visz	4
8	elme – szerelem	4
9	is – mégis	4
10	nyom – ottan	4
11	ellen – szellem	4
12	egyre – végtelen	4
13	ragyog – van	4
14	vágy – ágy	4
15	szerelem – én	4
16	költemény – én	4

19. táblázat. Rímpárok gyakorisági listája tokenek és lemmák alapján – a rímpárok tagjainak a sorrendje nem számít.

A 20. táblázatban a rímpárok tíz leggyakoribb szófaj- és szótagszám-kombinációját tüntettem fel a rímpárok tagjainak sorrendjét figyelmen kívül hagyva (vagyis a főnév-ige vagy a szótagszámra vonatkozó 2-3 listaelemek az ige-főnév és a 3-2 kombinációkra is vonatkoznak).

Szófaj

1	főnév – főnév	1312
2	főnév – ige	1225
3	ige – ige	608
4	melléknév – főnév	372
5	határozószó – főnév	331
6	főnév – névmás	236
7	melléknév – ige	189
8	határozószó – ige	153
9	névmás – ige	126
10	névutó – főnév	109

Szótagszám

1	2 – 3	1144
2	2 – 2	1088
3	2 – 4	601
4	3 – 3	590
5	1 – 2	377
6	3 – 4	364

7	1 – 3	292
8	4 – 4	192
9	1 – 1	177
10	2 – 5	116

20. táblázat. Rímpárok gyakorisági listája szófajok és szótagszámok alapján – a rímpárok tagjainak a sorrendje nem számít.

## 7. Összegzés

A tanulmány József Attila verseinek a példáján keresztül mutatta be a vershangzáshoz kapcsolódó jellemzők automatikus feltárásának egy módját, amely két fő lépésből állt. Az első lépés a korpusz automatikus annotálása volt. Ennek során egy XQuery szkripttel annotáltam a versek szerkezeti egységeit, azaz a versszakokat és a sorokat, valamint az *e-magyar* és a *hunpoem\_analyzer-TEI* program lefuttatásával a korpusz szavainak a grammatikai tulajdonságait és a vershangzáshoz kapcsolódó jellemzőket. A grammatikai tulajdonságok annotálása a szavak lemmájának, szófájának és morfoszintaktikai tulajdonságainak az automatikus elemzését jelentette. A vershangzás tulajdonságainak az annotálása pedig a szavak főbb fonológiai tulajdonságainak, a versszakok rímkepletének, a rímpároknak, a sorok ritmusának, valamint az alliterációknak az automatikus elemzésére terjedt ki. A korpusz hangzásjellemzőinek feltárásában a második lépés a gyakorisági listák generálása volt, amelyhez az XML-adatbázisok lekérdezésére szolgáló XQuery nyelvet használtam. A vershangzás automatikus elemzése számos ponton továbbfejleszthető. A távlati tervek között szerepel az időmértékes ritmusból metrumra absztrahálás funkciójának, az ütemhangsúlyos verselés megállapításának, valamint a rímek fokozati alapon történő elkülönítésének a beépítése az annotáló programba.

Nem volt célom, hogy a gyakorisági listák alapján következtetéseket vonjak le József Attila költészetéről. Csupán annak bemutatására törekedtem, hogy hogyan lehet a vershangzás jellemzőire vonatkozó olyan kvantitatív adatokhoz jutni, amelyek alapját adhatják különböző irodalomtudományos vizsgálatoknak. Az ilyen vizsgálatok nem pusztán egy adott költő verseire terjedhetnek ki, hiszen különböző alkotók verseiből nyert kvantitatív jellemzők összevetéséből derülhet ki, hogy mi az, amiben sajátos, és mi az, amiben általános mintázatokat valósít meg egy életmű vagy annak egy része. A lexikai tulajdonságok és a vershangzás jellemzőinek az automatikus feltárása arra is lehetőséget ad, hogy egy-egy időszakra, vagy éppen egy adott lírai műfajra vonatkozóan állapítsunk meg kvantitatív jellemzőket. Irodalmi szövegek kvantitatív jellemzőinek a feltárása, illetve az ilyen jellemzők alapján történő vizsgálatok természetesen nem helyettesíthetik a szövegek szoros olvasásával megvalósuló kvalitatív elemzéseket. A két megközelítést egymást támogató módszerként érdemes kezelni, amelyek más-más perspektívából tekintenek az irodalmi szövegekre, és így eltérő típusú kérdésekre adhatnak válaszokat.

## Automatic Analysis of Sound Devices in Attila József's Poems

The paper presents a method of automatic analysis of sound devices by using Attila József's poems as a case study. The first half of the paper discusses the functions of the program *hunpoem\_analyzer-TEI*, which was developed for the automatic annotation of sound devices and addresses the main steps of the annotation process. The second half of the paper presents different frequency lists of lexical features and sound devices extracted from the annotated corpus.

Keywords:

automatic poetry analysis, corpus linguistics, sound devices, rhyme, rhythm, Attila József, *hunpoem\_analyzer-TEI*, *e-magyar*





## Harangozó Ádám

*független kutató*

adamhrngzo@gmail.com

# FactGrid: adatbázis történészeknek

A FactGrid egy szabadon felhasználható adatokkal dolgozó, közösségi adatbázis történészeknek. A könnyen szerkeszthető, Wikibase alapú platformon a különböző kutatási projektek egy adatbázisban tárolhatnak és elemezhetnek strukturált adatokat. Az összekapcsolt adatok hosszútávon szerkeszthetőek maradnak, továbbá komplex lekérdezéseket végezhetünk rajtuk, áttekinthetünk hálózati kapcsolatokat, illetve megjeleníthetjük az adatokat térképen vagy idővonalon is. Tanulmányomban a platform bemutatásán túl azt igyekszem feltérképezni, mivel járulhat hozzá a hazai tudományhoz ez a szabadon felhasználható adatokkal dolgozó kezdeményezés.

Kulcsszavak:

adatbázis, FactGrid, Wikibase, strukturált adatok, összekapcsolt nyílt adatok, történettudomány



## 1. Bevezetés

Már magyarul is beszél a FactGrid, egy közösségi adatbázis történészeknek.<sup>1</sup> A közösségi jelző semmiképp sem a szakértelem hiányát jelöli – az adatbázist kizárólag kutatók szerkeszthetik –, hanem a szemléletet. Tanulmányomban a platform bemutatásán túl azt igyekszem feltérképezni, mivel járulhat hozzá a hazai kutatásokhoz ez a szabadon felhasználható adatokkal dolgozó kezdeményezés.

A platform ötlete az Erfurti Egyetemen végzett kutatásból született,<sup>2</sup> ahol a 18. századi Illuminátus-rend dokumentumaiból építettek adatbázist, feltérképezve a szervezet kapcsolati hálóját, szokásait, helyszíneit. Bár 2016-ban a projekt finanszírozása a Gotha Kutatóközpont által lezárult, egyre több, a témát kutató történész csatlakozott óriási adatgyűjtésekkel, az adatbázisszoftver és kiszolgálófelület pedig hamar elégtelennek bizonyult a növekvő felhasználói igényeknek. A komplex adatbázis követelményeinek jobban megfelelő, rugalmas, többnyelvű szoftvert keresve szövetkeztek a Wikimédia németországi szervezetével, hogy 2018-ban az általuk fejlesztett szoftverre építsék fel az új platformot, a FactGridet. Az új oldal már azzal a céllal készült, hogy egy általános nemzetközi adatbázis és kutatói környezet legyen történészek számára.

<sup>1</sup> „FactGrid: Kezdőlap,” *FactGrid*, hozzáférés: 2020.06.12, <https://database.factgrid.de/wiki/Kezdőlap>.

<sup>2</sup> „FactGrid: Gotha Illuminati Research Base Team,” *FactGrid*, hozzáférés: 2020.06.12, [https://database.factgrid.de/wiki/FactGrid:Gotha\\_Illuminati\\_Research\\_Base\\_Team](https://database.factgrid.de/wiki/FactGrid:Gotha_Illuminati_Research_Base_Team).

## 2. Wikibase, Wikidata, FactGrid

A platform lelke a nyílt forráskódú *Wikibase*.<sup>3</sup> Ezt a szoftvert a Wikimédia fejlesztette a Wikidata<sup>4</sup> számára, amely strukturált, központi adatbázisként szolgál a világ összes Wikipédiájának, ezen felül pedig bárki szabadon használhatja. A Wikidata segítségével a Wikipédián egy téma bármely nyelven íródott szócikke egyazon elemre mutat az adatbázisban.<sup>5</sup> Ezek az elemek nemcsak a különféle nyelvi verziókat kötik össze, de egy nagy ívű ontológiában<sup>6</sup> kapcsolódnak egymáshoz, amely a Wikipédiát is segít strukturálni.

Mielőtt rátérünk a FactGrid működésére, fontos leszögezni, hogy a történészi platform és a Wikidata nem vetélytársak vagy alternatívák, hanem egymást kiegészítő rendszerek. A Wikidata hasznosságát egyre több intézmény fedezi fel, és vannak kutatások, amelyek számára tökéletes otthont nyújt,<sup>7</sup> de az oldal két alapvető irányelve a legtöbb kutatási projektet kizárja. Ezek egyike a nevezetesség elve,<sup>8</sup> amely bár kevésbé szigorúan, mint a Wikipédián, de megköveteli, hogy minden létrehozott adatbáziselem valamiféle relevanciával rendelkezzen, például azáltal, hogy több egymástól független forrás is említi. A másik irányelv szerint a Wikidata nem adhat helyet saját kutatásunk első közlésének.<sup>9</sup> Ezek a Wikidata szempontjából nagyon is helyes elvek nem teszik lehetővé, hogy például komplett történelmi adathalmazokat töltsünk fel, tele eddig ismeretlen vagy akár azonosíthatatlan személyekkel.

Ezt a hiányt pótlandó, a FactGrid kiválóan használható nemcsak mint hosszútávú tárhely, hanem mint munkafelület, adatok elemzéséhez, vizualizációjához. Bármilyen új vagy már lezárult kutatás során keletkezett adathalmazt feltölthetünk, legyen az egy történelmi címjegyzék összes személye, folyóiratok előfizetői, klubok tagjai és így tovább.

<sup>3</sup> A szoftver bemutatását lásd: „Wikibase,” hozzáférés: 2020.06.12, <https://wikiba.se/>.

<sup>4</sup> Wikidata, „Wikidata: Kezdőlap,” hozzáférés: 2020.06.12, <https://www.wikidata.org/wiki/Wikidata:Kezdőlap>.

<sup>5</sup> Például: Wikidata, „Hunyadi János (Q242688),” hozzáférés: 2020.06.12, <https://www.wikidata.org/wiki/Q242688>.

<sup>6</sup> Wikidata, „Wikidata Generic Tree,” hozzáférés: 2020.06.12, <https://wikidata-todo.toolforge.org/tree.html?q=Q35120&rp=279&lang=en&method=list&depth=2>.

<sup>7</sup> Ilyen például a COURAGE-projekt (Kulturális ellenállás. Az ellenzékiesség öröksége az egykori szocialista országokban), hozzáférés: 2020.06.12, <http://hu.cultural-opposition.eu/registry/>. Erről bővebben lásd: Ghazal Faraj and András Micsik, „Enriching Wikidata with Cultural Heritage Data from the COURAGE Project,” Emmanouel Garoufallou, Francesca Fallucchi and Ernesto William De Luca, eds., *Metadata and Semantic Research: 13th International Conference, MTSR 2019, Rome, Italy, October 28–31, 2019, Revised Selected Papers*, Communications in Computer and Information Science book series 1057 (Cham: Springer, 2019), 407–418, [https://doi.org/10.1007/978-3-030-36599-8\\_37](https://doi.org/10.1007/978-3-030-36599-8_37).

<sup>8</sup> „Wikipédia: Nevezetesség,” *Wikipédia*, hozzáférés: 2020.06.12, <https://hu.wikipedia.org/wiki/Wikipédia:Nevezetesség>.

<sup>9</sup> „Wikipédia: A Wikipédia nem az első közlés helye,” *Wikipédia*, hozzáférés: 2020.06.12, [https://hu.wikipedia.org/wiki/Wikipédia:A\\_Wikipédia\\_nem\\_az\\_első\\_közlés\\_helye](https://hu.wikipedia.org/wiki/Wikipédia:A_Wikipédia_nem_az_első_közlés_helye).



### 3. Hogyan működik?

Az adatbázis felülete könnyen átlátható, egyszerűen szerkeszthető. Minimális programozói tudásra csak a lekérdezéseknél lehet szükségünk (SPARQL), de ehhez is rendelkezésünkre állnak vizuális alternatívák.

Ahogy minden Wikibase alapú adatbázis, a FactGrid is egyedi azonosítóval, Q-számmal ellátott elemekből áll, azaz Q147798 = Johann Sebastian Bach. A numerikus azonosító mellett minden elem címkékkel rendelkezik, amelyeket bármilyen nyelven megadhatunk, illetve az ezekhez külön-külön kapcsolódó leírásokkal is. Fontos megjegyezni, hogy a FactGrid (és a Wikidata) Q-számai nemcsak egyedi, de állandó azonosítók is. Sosem kerülnek újrafelhasználásra, duplikált elemek összevonása után pedig a kiürített azonosító átirányító oldalként funkcionál azon elem felé, amelyben összevonásra került, tehát a régi elemre mutató külső linkek sem évülnek el.

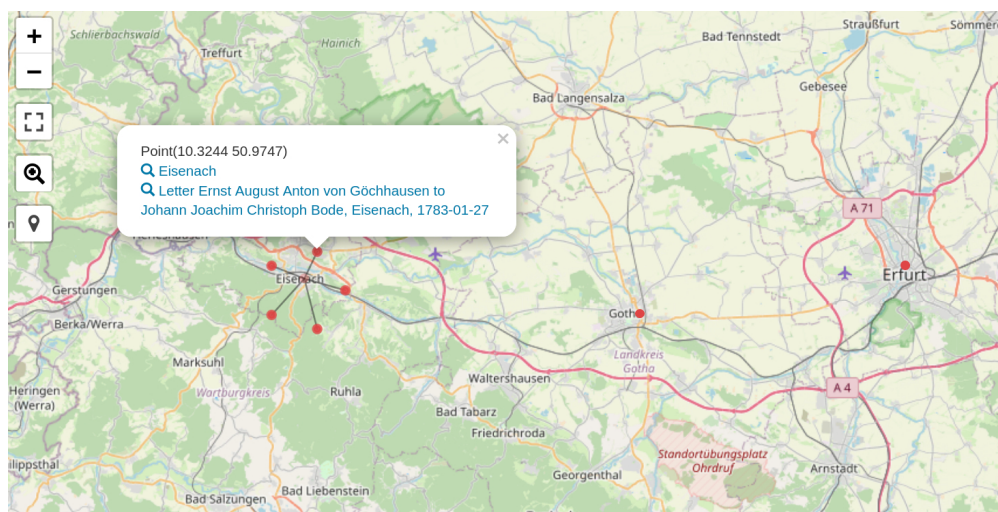
Az elemeket állításokkal írhatjuk le, amelyek tulajdonságok és értékek párosából állnak. A tulajdonságok egy saját számrendszerben mozognak „P” előjellel, a hozzájuk kapcsolt érték pedig leggyakrabban szintén egy adatbáziselem (Q), de lehet egyszerű szöveg, link vagy médiafájl is. Vagyis: Johann Sebastian Bach (Q147798) gyermeke (P150) Johann Christian Bach (Q147795). Egy tulajdonság több értékkel is rendelkezhet, esetünkben, mivel népes a Bach család: Johann Sebastian Bach (Q147798) gyermekei (P150) Carl Philipp Emanuel Bach (Q165366), Wilhelm Friedemann Bach (Q165365), Johann Christian Bach (Q147795), stb. A numerikus azonosítókat nem szükséges fejből tudnunk, elég elkezdni begépelni az adott címkét a saját nyelvünkön, és a rendszer felajánlja a megfelelő elemet. Minden elem nagyobb kategóriákba sorolódik (Prága → város → közigazgatási egység), ahogy a tulajdonságok is (feleség → házastárs → családi kapcsolatok → emberi kapcsolatok → elemek közötti kapcsolatok).

A P-számok egy csoportja minősítőként használható, tehát állításokat pontosíthatunk, árnyalhatunk velük. Rákosi Mátyás (Q165367) házastársa (P84) Feodora Fjodorovna Kornijlova (Q165368) + házasság kezdete (P49) 1942 (dátum).

lakhely	 <b>Gotha</b> <i>English</i>	 szerkesztés
	kezdő dátum	20. május 1783 <i>Gergely</i>
	befejező dátum	21. április 1785 <i>Gergely</i>
	▼ 1 hivatkozás	
	elsődleges forrás	Letter Friedrich Christian Rudorf to Johann Joachim Christoph Bode, Gotha, 1785-04-14 <i>English</i>
		+ forrás hozzáadása
	 <b>Buttstädt</b> <i>English</i>	 szerkesztés
	kezdő dátum	22. április 1785 <i>Gergely</i>
	megjegyzés	passes through Weimar on the 22nd
	▼ 1 hivatkozás	
	elsődleges forrás	Friedrich Christian Rudorf, Quibus Licet, Buttstädt, 1785-05-24 <i>English</i>
		+ forrás hozzáadása

1. ábra. Friedrich Christian Rudorf adatlapja (részlet)

Minden egyes állítást alátámaszthatunk forrásokkal, akár többel is. A FactGriden megadhatunk egymással versengő információkat is, mindegyiket saját hivatkozásokkal, például ha esemény időpontja különböző forrásokban eltérő dátumok alatt szerepel. Ezt nemcsak akkor tehetjük meg, ha nem tudjuk, melyik áll közelebb az igazsághoz, hanem akkor is, ha szeretnénk felsorolni a különböző variánsokat, modellezni egy történeti vitát vagy jelezni szeretnénk egy gyakran előforduló tévedést. A különböző információkat egyszerűen rangsorolhatjuk, kiemelhetjük az aktuálisan elfogadott értéket, illetve minősítővel jelezhetjük a megbízhatóságukat. Ez a rugalmasság annak is teret enged, hogy létrehozzunk új elemeket vagy elemek közötti kapcsolatokat pusztán kísérleti célból, például egy feltevés vagy munkahipotézis teszteléséhez.<sup>10</sup>



2. ábra. Illuminátus-dokumentumok származási helyei térképre vetítve

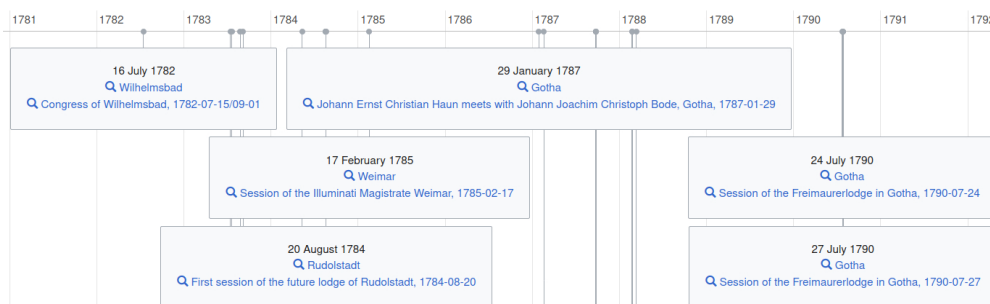
Az FactGrid-elemeket a belső kapcsolatokon felül külső adatbázisokhoz is hozzákapcsolhatjuk. Bachnál például megadhatjuk, milyen azonosítóval szerepel a GND-ben, a Nemzeti Névtérben, a VIAF-ban, a Bach Digital-projektben, illetve bármilyen egyéb oldalon, katalógusban, amelyekre szükségünk van a kutatásunkhoz. Ugyanígy tulajdonságokat is megfeleltethetünk más ontológiákban elfoglalt helyüknek, ezáltal „hozzátérképezhetjük” az adatmodellünket másokéhoz. Külső adatbázisok vissza is utalhatnak a FactGridre, a Wikidatára például már több könyvtári és múzeumi katalógus, köztük a VIAF<sup>11</sup> és a MoMa<sup>12</sup> is hivatkozik. Bár még nem alakult ki gyakorlata, a releváns Q-számokat elhelyezhetjük cikkekben, tanulmányokban, akár blogbejegyzésekben is, illetve bárhol hivatkozhatunk velük például a projektünk során létrehozott adatkészletre.

<sup>10</sup> A működési elvek itteninél részletesebb kifejtése a projekt blogján magyarul is olvasható, illetve egy kis kalandozással az oldalon vagy a Wikidatán könnyű és élmény kiismerni azokat. Lásd Olaf Simons, „FactGrid GYIK – Miért használjam a FactGridet a kutatási projektemhez?,” <https://blog.factgrid.de/archives/1955>.

<sup>11</sup> Például VIAF, „Arany János (95251869),” hozzáférés: 2020.06.12, <https://viaf.org/viaf/95251869/>.

<sup>12</sup> Például MoMa, „Marina Abramović (26439),” hozzáférés: 2020.06.12, <https://www.moma.org/artists/26439>.

Az adatbázisra egy hagyományos Wikipédia-külsejű (MediaWiki) réteg is került, annak minden funkciójával. Itt szabadon létrehozhatunk az adatainkkal összekapcsolt „szócikkeket”, ahol közölhetünk szövegeket, átiratokat vagy készíthetünk külön oldalt projektünknek, ahol a kutatócsoport koordinációját végezhetjük.<sup>13</sup>



3. ábra. Idővonal: események, amelyeken J. J. C. Bode részt vett

#### 4. Bábeli rend

A FactGrid egyik alapvető funkciója, hogy bármilyen nyelven „írhatjuk”, a felhasználó pedig saját nyelvén olvashatja: tehát nemzetközi. Mivel a rendszer maga csak Q-számokkal dolgozik, ezért amikor megadjuk például a Bach-foglalkozása-zeneszerző hármast saját nyelvünkön, a többi felhasználó ugyanazokat a Q- és P-számokat látja, mindössze saját nyelvén felcímkezve. Ezáltal, ha a világ másik pontján javítanak, bővítenek egy számunkra fontos adaton, az egyből felhasználhatóvá válik anyanyelvünkön is. A nyelvi korlátok nélküli szabadság a FactGrid segédeszközeivel készíthető vizualizációkban, idővonalakon, térképeken is megjelenik. Ehhez a többnyelvűséghez viszont szükséges, hogy az elemek címkéi létezzenek az adott nyelven. A mindenki számára nyitott és egyszerű fordítási folyamatot is felgyorsítja a közösségi szerkesztés, minél több a felhasználó, annál jobban.

Az oldalon található projektek többsége egyelőre nem kifejezetten kutatás, inkább előzetes adatbázis-építés és a platform nyújtotta lehetőségekkel való kísérletezés. Egyikük például a Szász–Gotha–Altenburg Hercegség területén lévő plébániák és lelkészek hálózatát térképezi fel 1500 és 1920 között.<sup>14</sup> A későbbi kutatás megalapozásához a Thüringer Pfarrerbuch-sorozat első kötetét<sup>15</sup> alakították FactGrid-adatokká: a könyv felsorolja a Hercegség összes lelkészét és a plébániák egyéb dolgozóit, illetve életrajzaikat, leszármazásukat. A legnagyobb kihívást az jelentette, hogy minden személy saját, külön azonosítót kapjon, mivel a nevek gyakran öröklődtek családon belül,

<sup>13</sup> Például „FactGrid: The Global Genealogy of Lodges,” *FactGrid*, hozzáférés: 2020.06.12, [https://database.factgrid.de/wiki/FactGrid:The\\_Global\\_Genealogy\\_of\\_Lodges](https://database.factgrid.de/wiki/FactGrid:The_Global_Genealogy_of_Lodges).

<sup>14</sup> „FactGrid: Religion im Herzogtum Gotha-Sachsen-Altenburg,” *FactGrid*, hozzáférés: 2020.06.12, [https://database.factgrid.de/wiki/FactGrid:Religion\\_im\\_Herzogtum\\_Gotha-Sachsen-Altenburg](https://database.factgrid.de/wiki/FactGrid:Religion_im_Herzogtum_Gotha-Sachsen-Altenburg).

<sup>15</sup> Bernhard Möller, hrsg., *Thüringer Pfarrerbuch. Band. 1: Herzogtum Gotha. Schriftenreihe der Stiftung Stoye der Arbeitsgemeinschaft für Mitteldeutsche Familienforschung e.V.*, Band 26 (Neustadt an der Aisch: Verlag Degener & Co., 1995).

és a kötetben szereplő életrajzok nem egyértelműsítik, pontosan kire utalnak. A duplikátumok nagy részét más forrásokkal összevetve sikerült kiküszöbölni, így jelenleg 13 344 személyből áll az adathalmaz, részletes állításokkal azok családi kapcsolatairól, a plébániákban betöltött szerepeikről és ezek időpontjairól. A lelkési családfák és plébániák térképre vetítésével az adatbázis lehetővé teszi, hogy a kutatók olyan kérdéseket vizsgálhassanak, mint hogy milyen előnyökkel szolgálhattak a házasságkötések a karrier szempontjából, hogyan változott az idők során a családi befolyás a gyermekek egyházi pályaválasztására, vagy mennyiben nőtt túl a helyi lelkészhálózatokon az egyházi kinevezettek földrajzi elhelyezkedése. A projekt egyik további célkitűzése annak vizsgálata, hogyan lehetne használni a FactGridet a könyvsorozat jövőbeni köteteinek előkészítéséhez.<sup>16</sup>

 <a href="#">wd:Q41462</a>	Christian August Härter	* 1750-12-24 Aschara, + 1815-04-01 Gräfenonna	24 December 1750	 <a href="#">wd:Q43514</a>	Aschara	1 April 1815	 <a href="#">wd:Q43543</a>	Gräfenonna
 <a href="#">wd:Q41463</a>	Christian August Ludwig	* 18 November 1701, + 21 November 1762, pastor in Bienstädt and Goldbach	18 November 1701	 <a href="#">wd:Q77181</a>	Zeitz	21 November 1762	 <a href="#">wd:Q36236</a>	Goldbach
 <a href="#">wd:Q41464</a>	Christian August Thienemann	* 1717-04-07 Etzdorf (Thüringen), + 1780-07-30 Orlamünde, 1744 field preacher in Gotha	7 April 1717	 <a href="#">wd:Q76985</a>	Etzdorf (Thuringia)	30 July 1780	 <a href="#">wd:Q77099</a>	Orlamünde

4. ábra. A Thüringer Pfarrerbuch adatbázisként: személyek, születési és halálozási helyek, idők

Az első nemzetközi kutatás a FactGriden a korai gyarmati Mexikóval foglalkozik. A Lancsteri Egyetem digitális bölcsészeti központja, a mexikóvárosi Templo Mayor Múzeum és a Lisszaboni Egyetem számítógép-tudományi kutatóközpontja az Új-Spanyolországból küldött jelentéseket veti alá számítógépes elemzésnek, azt kutatva, hogyan jellemezték a spanyol gyarmatosítók az őslakosokat és földjeiket, illetve hogyan használták ezeket az információkat saját céljaikra.<sup>17</sup>

Fontos megjegyezni, hogy a FactGridnek, bár egyre többen használják, még vannak hiányosságai. A fordítás még csak kevés nyelven megkezdett, magyarul közel sem teljes, illetve az oldal egyelőre nem rendelkezik alapos útmutatókkal. Az adatmodell sem mindenre kiterjedő, például továbbra is kérdéses, külön elemekként vagy egy folytonosságként szerepeljenek a bezárt majd újraindított intézmények. Mindenesetre a közösségi modell miatt a rendszer fejlesztésére irányuló kezdeményezések könnyebben lendületet kaphatnak, és ebben minden felhasználó részt vehet, jelentse ez egy-egy címke lefordítását vagy új funkciók fejlesztését.

## 5. Az adat mint közkinccs

A mind jobban neoliberais, piaci keretbe szorított egyetemek oktatóin folyamatos a nyomás, hogy minél többet publikáljanak, másszák a ranglétrát, hogy minden kutatás

<sup>16</sup> Olaf Simons, „The First Volume of the Thuringian Pastor’s Book (1500–1920) as a Wikibase Data Set,” <https://blog.factgrid.de/archives/1941>.

<sup>17</sup> „FactGrid: Digging into Early Colonial Mexico,” *FactGrid*, hozzáférés: 2020.06.12, [https://database.factgrid.de/wiki/FactGrid:Digging\\_into\\_Early\\_Colonial\\_Mexico](https://database.factgrid.de/wiki/FactGrid:Digging_into_Early_Colonial_Mexico).



sikerrel, hasznosítható eredménnyel záruljon. Mivel a legtöbb – ha nem minden – kutató ebben a tudástermelési rendszerben kényszerül dolgozni, joggal merülhet fel a kérdés: miért ossza meg a kutató közkincként a komoly erőfeszítések árán kibányászott adathalmazokat? A FactGrid használatával azonban nem veszik el a kutatói én.

Minden FactGrid-adat CC0-licenccel<sup>18</sup> rendelkezik, más szóval kutatásunk közkincként lesz elérhető és szabadon felhasználható. A CC0 nem kötelez forrásmegjelölésre, ez azonban nem jelenti azt, hogy a felhasználók nem is fognak majd hivatkozni a szerzőre. A FactGrid (és a legtöbb nyílt licencű projekt) kifejezetten ösztönzi a felhasznált adatok forrásmegjelölését, mindössze nem jogi ostonnal, hanem kommunikációval és a szerzőiség átláthatóságával. A FactGriden minden egyes szerkesztés, legyen az egy apró korrekció vagy ezernyi új adat, a felhasználó saját fiókjához kapcsolódik, tehát bármit is csinálunk, az automatikusan az aláírásunkkal rendelkezik. Bizonyítani sem körülményes, hisz szerkesztési történetünk mindenki számára nyilvános.

Ellenben, ha kifejezetten hangsúlyozni szeretnénk factgrides munkásságunkat, létrehozhatunk olyan gyűjtőoldalakat is, amelyek az összes szerkesztésünket tartalmazzák,<sup>19</sup> illetve saját adataleltárunk is részletes önéletrajzként szolgálhat.<sup>20</sup> A szerkesztési összefoglalókat felhasználhatjuk a projekt beszámolójában, szakmai önéletrajzunkban vagy akár afféle minitanulmányként is. Nemcsak hogy szerzői leszünk adatainknak és modelljeinknek, de ezek, jogi korlátozások nélkül, könnyen túlszárnyalhatják saját projektünket, és más kutatások, publikációk számára is értékes forrássá válhatnak, mindvégig a saját nevünk alatt. Bár adatbázisok esetén nagy számokról beszélünk, a közösségi szerkeszthetőség révén egy módosítás, egy új tulajdonság vagy kapcsolat is jelentős változást hozhat.

Az átláthatóság és a széles körű szerkeszthetőség az elévülés legjobb ellenszere is. Az internetes közlés régóta ismert előnye a nyomtatott betű rögzítettségével szemben, hogy online utólag is javíthatjuk, bővíthetjük munkánkat. Ugyanakkor, ha egy projekt már rég lezárult, nem finanszírozott, nem profitál, gyakran nincs lehetőségünk a hibák javítására vagy frissíteni új eredményekkel, talán még szerkesztői hozzáférésünk sincs. Ha adatainkat egy nyílt rendszerbe visszük fel, nemcsak általunk lesznek javíthatóak, hanem amikor mi már egyéb projektekkel foglalkozunk, mások is frissíthetik őket. Így például nem történhet meg, hogy néhány elkövetett hiba vagy részbeni elévülés miatt kutatásunk elfelejtődik.

A FactGrid, bár ugyanúgy hasznos és működőképes a teljesítmény alapú, a metrika bővületében felgyorsuló egyetemi rendszerben, igazából egy értékeken nyugvó, emberközpontú, lassú<sup>21</sup> tudomány keretein belül értelmezhető. Egyszerre lehet modern, nemzetközi szinten felmutatható eszköz a humántudományok elértéktelenődésével

<sup>18</sup> „CC0,” Creative Commons, hozzáférés: 2020.06.12, <https://creativecommons.org/share-your-work/public-domain/cc0/>.

<sup>19</sup> Mivel lekövethető minden változás szerzője, időpontja és tartalma, akár meta-történetírásként, historiográfiaként is tekinthetünk rá.

<sup>20</sup> Például *FactGrid*, „Olaf Simons (Q11298),” hozzáférés: 2020.06.12, <https://database.factgrid.de/wiki/Item:Q11298>.

<sup>21</sup> Alison Mountz, Anne Bonds, Becky Mansfield, Jenna Loyd, Jennifer Hyndman, Margaret Walton-Roberts, Ranu Basu, Risa Whitson, Roberta Hawkins, Trina Hamilton and Winifred Curran, „For Slow Scholarship: A Feminist Politics of Resistance through Collective Action in the Neoliberal

szemben, és strukturális elmozdulás a nyílt tudomány felé. Higgadt, mégis radikális kezdeményezés a digitális humanizmus bécsi kiáltványa,<sup>22</sup> sőt a barlowi függetlenségi nyilatkozat<sup>23</sup> szellemében. Használata lehetőséget nyújt tapasztalt kutatók számára, hogy reflektálhassanak a jelenlegi rendszerre és abban elfoglalt helyükre, illetve egyetemi hallgatóknak, hogy egy átlátható, nem tekintélyelvű rendszerben válhassanak kutatókká.

## 6. A Wiki-galaxis felé

Miközben az egyetemi folklór része a Wikipédia-használat tiltottsága,<sup>24</sup> egyre több oktató fedezi fel, hogy az oldal irányelvei valójában igen hasonlóak a tudományos szövegek követelményeihez, sőt a Wikipédia-szerkesztés fejleszti a hallgatók írás- és kutatási készségeit.<sup>25</sup> Emiatt számos egyetem kezdte beemelni a tananyagba a Wikipédia-szerkesztést, például szemináriumi feladatok szintjén.<sup>26</sup>

Ennek mintájára megfontolandó a FactGrid alkalmazása az egyetemi oktatásban is: használatával a diákok a jövőben elengedhetetlen digitális bölcsészeti szoftvereket tanulhatnak meg kezelni, köztük a Wikipédiát működtető MediaWikit is, és elsajátíthatják a közösségi munkát, illetve azt a szemléletet, amely a tudást és a kultúrát közkinsnek tekinti. Mindeközben pedig változatlanul tudományos környezetben maradnak.

A FactGrid kiváló lehetőség, hogy az oktatók együtt dolgozhassanak a hallgatókkal, méghozzá nem egy szimulált, ingerszegény, hanem egy olyan környezetben, ahol, jobb esetben, a tanár maga is dolgozik. Így a diákok korán bevonódhatnak a kutatói munkába, méghozzá nem névtelenül elvégzett munkával, hanem olyasmivel, amelynél súlya van minden hozzájárulásuknak. Egy szeminárium esetén, például a közös platformot szerkesztve, megtanulhatnak együttműködni, segíthetik, inspirálhatják egymást. Közben pedig mind a tanár, mind a diák folyamatosan (át)láthatja a másik munkáját, módszereit.

Végső soron a FactGrid használata átjárást, élő kapcsolatot teremt a Wikidata és hosszabb távon a Wikipédia felé. A platformon keresztül megismerkedni a Wikidatával a kutatók számára sem csak közszolgálat: amellett, hogy könnyen áttölthetjük a Wikidata számára releváns adatokat, az oldal többmillió, tízezrek által karbantartott állományából és adatmodelljéből szilárd alapot generálhatunk kutatásunknak.

---

University,” *ACME: An International Journal for Critical Geographies* 14 (2015): 1235–1259, <https://www.acme-journal.org/index.php/acme/article/view/1058>.

<sup>22</sup> Hannes Werthner et al., „Vienna Manifesto on Digital Humanism,” hozzáférés: 2020.06.12, <https://www.informatik.tuwien.ac.at/dighum/index.php>.

<sup>23</sup> John Perry Barlow, „Cyberspace Függetlenségi Nyilatkozat,” *Replika* 26 (1997): 165–166, [http://replika.hu/system/files/archivum/replika\\_26\\_14\\_barlow.pdf](http://replika.hu/system/files/archivum/replika_26_14_barlow.pdf).

<sup>24</sup> Érdemes viszont feltenni magunknak a kérdést: ki szerkesztheti a Wikipédiát?

<sup>25</sup> Matthew A. Vetter, Zachary J. McDowell and Mahala Stewart, „From Opportunities to Outcomes: The Wikipedia-Based Writing Assignment,” *Computers and Composition* 52 (2019): 53–64, <https://doi.org/10.1016/j.compcom.2019.01.008>.

<sup>26</sup> Corrado Petrucco, „Wikipedia in University Courses: Teaching Practices and Educational Benefits,” *Research on Education and Media* 10, 2 (2018): 10–16, <https://doi.org/10.1515/rem-2018-0010>.



## 7. Hozzáférés

Érthető tehát, miért gyümölcsöző a platform használata a kutatói közösségek számára. A FactGrid a kutatási innováció mellett talán akkor nyújthatja a legtöbbet, ha elősegíti az akadémiai világ mélyebb önreflexióját, amely túlmutat a hozzáférés kérdéskörén. Bár még ma is távolinak tűnhet a tudományos közösséget átható, a szakmai publikációkon túlmutató nyílt hozzáférés-szemlélet, fel kell tennünk már most a kérdést: mennyit ér a hozzáférés valójában, ha a szélesebb közönség, sőt talán a kutatók jelentős része sem tud az adott szöveg létezéséről? Horváth Iván 1999-ben rendkívüli kiállítást rendezett, ahol elveszett magyar honlapok példájával figyelmeztetett az internetes szövegek tárolásának problematikájára.<sup>27</sup> Külön kiállítást érne viszont az a számtalan, amúgy remek tudományos honlap, virtuális galéria, oktatási és ismeretterjesztő anyag, amely ma is elérhető, de a kommunikáció hiánya miatt csak szűkebb közönség tud a létezéséről, talán még a szakmában sem mindenki, akinek hasznára válhatna.<sup>28</sup> Ebben nemzetközi szinten sem maradunk el.<sup>29</sup> Az archiválás önmagában nem a használatot biztosítja, csak annyit, hogy a szűk kutatói közösség időről időre újra felfedezheti majd ezeket a szövegeket mint történelmi „tárgyakat”. Tudománykommunikáció, tájékoztatás nélkül ugyanis nem válnak a közös tudás részévé; sarkosan fogalmazva: amiről nem tud a közönség, az gyakorlatilag nem is létezik.<sup>30</sup> Az obskúrus, elfelejtett honlapok és kutatási projektek nem csak elmulasztott lehetőségek: a közpénzből létrehozott tudás kommunikációja, a megosztás a kutatók alapvető feladata, nemcsak az egyetem falain belül, hanem azon kívül is. Végző soron a tudomány mint társadalmi intézmény nem a belső buborékok felfújogatását szolgálja, hanem hídépítést az azon kívül eső közönséghez: saját határainak eltörlését. A FactGrid használata pedig egy lépést jelenthet ebbe az irányba.

### FactGrid: Database for Historians

FactGrid is an open data-based community database for historians. The platform, running on the easy-to-edit *Wikibase*, provides research projects a shared database where they can store and analyse structured data. The linked data remains editable in the long-term, we can run complex

<sup>27</sup> Horváth Iván, „Magyar tartalomipar,” <http://magyar-irodalom.elte.hu/contentware/magyar/>.

<sup>28</sup> Néhány kiragadott példa a sok közül: „Írónők a hálón 1880–1945,” <http://ironok.elte.hu/>; „A budai vár,” <https://budavar.btk.mta.hu/>; „József Attila összes műve,” <http://jozsefattila.elte.hu/>; „A magyar nyelv finnugor alapjai,” <http://finnugor.elte.hu/index.php?q=sztika>; „Hálózati tankönyvek, szöveggyűjtemények,” <http://magyar-irodalom.elte.hu/gepesk/tk.html>; „Tabán térképes képtár,” <https://web.archive.org/web/20110829000913/http://atlasz.tabanmuzeum.hu/>; Hadas Miklós, „Férfikutatások: Szöveggyűjtemény,” <http://publikaciok.lib.uni-corvinus.hu/publikus/643014.pdf>, hozzáférés 2020.06.12.

<sup>29</sup> „Rombase,” <http://rombase.uni-graz.at/>; *Online Encyclopedia of Mass Violence*, <https://www.sciencespo.fr/ceri/en/ouvrage/oemv>; *Lexicon of Scholarly Editing*, <https://lexiconse.uantwerpen.be/>; *Encyclopedia of Mediterranean Humanism*, <https://www.encyclopediae-humanisme.com/>; *Compendium heroicum*, <https://www.compendium-heroicum.de/>, hozzáférés 2020.06.12.

<sup>30</sup> Nehezítő tényező, hogy nincs még központi gyűjtőhelye sem a hasonló oldalaknak.

queries and visualise the data on map or timeline. In my article, besides introducing the platform, I will try to unpack the ways in which the initiative might enrich historical research in Hungary.

Keywords:

database, FactGrid, Wikibase, structured data, linked open data, history

<LABOR>



## Kiss Tamás

Monguz Információtechnológiai Kft./Qulto International

Eötvös Loránd Tudományegyetem Digitális Bölcsészeti Központ

tkiss@monguz.hu

## Kanász-Nagy Zoltán

Monguz Információtechnológiai Kft./Qulto International

zkanasznagy@monguz.hu

# Rekordoldali médiamegjelenítés a DSpace 7-ben

Közgyűjtemények és kutatói közösségek részéről az elmúlt évek során növekvő igény mutatkozott nyílt forráskódú, könnyen telepíthető, testre szabható és üzemeltethető repozitóriumi megoldásokra. A digitális objektumok tárolására, valamint repozitálására vonatkozó elvárások kielégítésére megfelelő megoldásokat nyújt a DuraSpace nevű, *not-for-profit* közösség által 2002 óta fejlesztett repozitóriumi szoftver, a *DSpace* jelenlegi, hetes számú verziója. A *DSpace 7* azonban egyelőre csak béta verzióban érhető el és olyan hiányosságokkal rendelkezik, amelyek megoldása az általunk kiszolgálni kívánt intézményi igények sürgőssége miatt nem várható a DuraSpace fejlesztői által meghatározott ütemterv szerinti implementációkra. Ezek közé tartozott a médiamegjelenítés és -lejátszás hiánya, melyek pótlására a cikk szerzői, valamint kollégáik (Tápai-Kovács Annamária webdizájnér és Sipos Dániel szoftverfejlesztő) vállalkoztak, felajánlva a létrejött kódot a DuraSpace közösség, és így a *DSpace 7* kódbázisa számára. A cikk a DuraSpace-szel megvalósult együttműködés rövid történetét tárgyalja és a vonatkozó fejlesztés médiatípusonkénti kódrészleteit adja közre.

Kulcsszavak:

DSpace 7, szoftverfejlesztés, médiamegjelenítés, médialejátszás, repozitórium, Angular 10



## A DSpace 7 közösségi fejlesztése

A *DSpace* egy nyílt forráskódú repozitóriumi szoftver, amely az elmúlt években nagy népszerűsége tett szert, elsősorban közoktatási és közgyűjteményi intézmények körében. Népszerűségét főként viszonylag egyszerű használata, konfigurálhatósága és

üzemeltethetősége révén nyerte el. Számos projekt valósult meg a *DSpace* repozitóriumi rendszernek köszönhetően, és ezen projektek tapasztalataiból merítve mára több kiegészítő (fizetős, illetve szabadon hozzáférhető *add-on*) készült el és vált hozzáadott értékkel az alapszoftver számára.

A *DSpace* hetedik verziója továbbra is a fejlesztői és felhasználói közösség ötleteire és igényeire alapulva készül,<sup>1</sup> de egyben új mérföldkő is a *DSpace* életében. Miközben a szoftver megörökölte az előző verziók hatékonynak bizonyult funkcióit, részben szakítva az eddigi verziók hagyományaival, teljesen új, webalapú, *Angular* 10 keretrendszerben megírt felhasználói felületet kapott.<sup>2</sup> Ezen felül több, *front-end* fejlesztésen túlmutató újdonság (pl. REST API és konfigurálható entitások) is belekerült abba a fejlesztési listába, amelyben a DuraSpace kijelölte az új verzió alapjait és a fejlesztés távlati céljait.

Tim Donohue, a *DSpace*-fejlesztés technológiai vezetőjének eredeti tervei szerint a 2018 óta tartó munka első kiadása (*release*) 2020 nyarán készült volna el. Előre nem látható erőforráshiány miatt azonban a fejlesztők a jelenlegi béta verzió elkészültének időpontját 2021 második negyedévére módosították, ezért a megjelenésre a felhasználóknak még néhány hónapot várniuk kell.<sup>3</sup>

Ez azonban nem jelenti azt, hogy a szoftver jelen állapotában nem használható: a végleges funkciók nagy része már rendelkezésre áll, így aki szeretné a *DSpace* 7-et az első *release* előtt saját vagy ügyfelei repozitálással kapcsolatos igényeinek kielégítésére használni, kipróbálhatja a szoftver hivatalos demóját,<sup>4</sup> vagy telepítheti a szoftver legutóbbi verzióját.

## Hozzájárulásunk a *DSpace* 7 fejlesztéséhez

Csapatunk az új *DSpace*-verzió életútját a 2020-as év első felében, az első, nagyközönség által is kipróbálható alfa verzió közzétételével egyidőben kezdte el követni.<sup>5</sup> Fejlesztőink először tesztelőként próbálták megismerni az új verzióban rejlő lehetőségeket, majd igyekeztek tapasztalataikról minél több visszajelzést adni a szoftver fejlesztői számára. Később elkezdtük az első verzió elkészültét kisebb fejlesztésekkel segíteni, melynek következtében tágabb szakmai „családunk”, a Monguz/Qulto International – immár a DuraSpace fejlesztői közösség tagjaként – közreműködött különböző, például a találati lista és a rekordoldal megjelenítését érintő hibák javításában.

<sup>1</sup> A *DSpace* szoftver fejlesztését szervezeti szinten a DuraSpace *not-for-profit* szervezet koordinálja, amely a működéséhez szükséges tőkét nem a szoftver eladásából, hanem a közösség tagjainak hozzájárulásából finanszírozza. A jelentősebb, a közösséget pénzügyileg is támogató szervezeti tagok névsora megtekinthető a következő weboldalon, hozzáférés: 2021.04.27, <https://duraspace.org/dspace/community/members/>.

<sup>2</sup> A fejlesztésben felhasznált *back-end* és *front-end* technológiák részletes leírása, hozzáférés: 2021.04.27, <https://wiki.lyrasis.org/display/DSDOC7x/Installing+DSpace>.

<sup>3</sup> Tim Donohue et al., „On the Road to *DSpace* 7: 2018 Edition”, 2018, [http://repositorio.concytec.gob.pe/bitstream/20.500.12390/46/1/2018\\_Donohue\\_Dspace7.pdf](http://repositorio.concytec.gob.pe/bitstream/20.500.12390/46/1/2018_Donohue_Dspace7.pdf).

<sup>4</sup> A demóverzió elérhető a következő linkre kattintva, hozzáférés: 2021.04.27, <https://dspace7-demo.atmire.com/home>.

<sup>5</sup> A DuraSpace közösség tagjai nemcsak a nagyobb léptékű *DSpace* Steering Group gyűléseken és a fejlesztői szintet érintő technikai státuszmegbeszéléseken kommunikálnak egymással, hanem a Slack közösségi chatcsatornán is, amelyre bárki szabadon regisztrálhat.

A DuraSpace ütemterve a képi, hang-, illetve videotartalmak közvetlenül webes felületen keresztül történő, rekordoldali megtekintését és lejátszását csak egy későbbi fejlesztési fázisban (legkorábban 2021 végén) tervezte lehetővé tenni a felhasználók számára. Csapatunk vállalta a rekordokhoz kapcsolódó médiafájlok megtekintését és lejátszását lehetővé tevő fejlesztési munkák elvégzését, ezzel is előrébb hozva a szoftver felhasználóbarát működésének megvalósulását. A fejlesztőink által implementált újdonság a következő (béta 5) verzióban már az alaprendszer részét fogja képezni. Akiket érdekel ennek a szoftver rendeltetésszerű használatát nagyban elősegítő fejlesztésnek a kronológiája, illetve a megvalósítás részletei, a következő Github-oldalakon bővebb információkhoz juthatnak: <https://github.com/DSpace/dspace-angular/issues/885>, <https://github.com/DSpace/dspace-angular/pull/888>.

## Publikáció: Csabai képek



### Leírás

Csabai képek kiállítása..

### Kulcsszavak

Plakát, kiállítás

### URI

<http://repository.qulto.eu/handle/12234>

### Állományok

rekord00000.jpg (5.26 MB)

### Kiadó

Békés Megyei Könyvtár

### Gyűjtemények

Szövegek, képek...

Teljes tárgy oldala

1. ábra. Videolejátszás megvalósítása (részlet a rekordoldalról)

## Médiamegjelenítések fejlesztési módszertana, kódrészletekkel

A képi, hang-, illetve videotartalmak rekordoldali megjelenítésének általunk használt fejlesztési módszertana és az ezekből származó kódrészletek fejlesztési fázisokra bontása a következő:

1. Rekordoldali médiamegjelenítés előkészítése: a *DSpace*-rekordhoz kapcsolódó összes *bitstream* leíró rekordját elemezve egy-egy *mediaViewerItem* típusú

objektumot hozzunk létre, azaz a *bitstream resource descriptort* átmappeljük *mediaViewerItem*mé:

```
src\app+item-page\media-viewer\media-viewer.component.ts

/**
 * This method loads all the Bitstreams and Thumbnails and converts
 * them to media item
 */

ngOnInit(): void {
  this.mediaList$ = new BehaviorSubject([]);
  this.isLoading = true;
  this.loadRemoteData('ORIGINAL').subscribe((bitstreamsRD) => {
    if (bitstreamsRD.payload.page.length === 0) {
      this.isLoading = false;
      this.mediaList$.next([]);
    } else {
      this.loadRemoteData('THUMBNAIL').subscribe((thumbnailsRD) => {
        for (
          let index = 0;
          index < bitstreamsRD.payload.page.length;
          index++
        ) {
          bitstreamsRD.payload.page[index].format
            .pipe(getFirstSucceededRemoteDataPayload())
            .subscribe((format) => {
              const current = this.mediaList$.getValue();
              const mediaItem = this.createMediaViewerItem(
                bitstreamsRD.payload.page[index],
                format,
                thumbnailsRD.payload && thumbnailsRD.payload.page[index]
              );
              this.mediaList$.next([...current, mediaItem]);
            });
        }
        this.isLoading = false;
      });
    }
  });
}
```

2. Képek megjelenítésére a Kolkov által javított *Ngx-gallery* modult használjuk.<sup>6</sup>  
A galéria működéséhez minden egyes képfájl esetén szükség van három külön-

<sup>6</sup> Andrey Kolkov, „Ngx-gallery,” 2020, hozzáférés: 2021. 04. 27, <https://www.npmjs.com/package/@kolkov/ngx-gallery>.



bőző képi változat meghatározására: egy kis és egy közepes méretű *thumbnail*re, illetve egy nagy méretű képre – ez utóbbi céljára az eredetileg feltöltött képfájl használjuk. Az alábbi kódrészben látható, hogy bélyegképként az eredeti képfájlhoz kapcsolódóan feltöltésre került *thumbnail* vagy, ha ilyen nem létezik, az eredeti képfájlból a *Dspace* által generált *thumbnail* kerül felhasználásra. Konkrét bélyegkép hiányában a fejlesztés biztosít egy általános felhasználású *thumbnailt* (*replacement\_image.svg*).

```
src\app\+item-page\media-viewer\media-viewer-image\
media-viewer-image.component.ts
```

```
/**
 * This method convert an array of MediaViewerItem into
 * NgxGalleryImage array
 * @param medias input NgxGalleryImage array
 */
convertToGalleryImage(medias: MediaViewerItem[]): NgxGalleryImage[] {
  const mappadImages = [];
  for (const image of medias) {
    if (image.format === 'image') {
      mappadImages.push({
        small: image.thumbnail
          ? image.thumbnail
          : './assets/images/replacement_image.svg',
        medium: image.thumbnail
          ? image.thumbnail
          : './assets/images/replacement_image.svg',
        big: image.bitstream._links.content.href,
      });
    }
  }
  return mappadImages;
}
```

3. Webes környezetben (*Angular framework*) a galéria komponens inicializálása, felhasználva a fent említett, Kolkov-féle *Ngx-gallery* modult:

```
src\app\+item-page\media-viewer\media-viewer-image\
media-viewer-image.component.html
```

```
<div [class.change-gallery]="isAuthenticated$ | async">
  <ngx-gallery
    class="ngx-gallery"
    [options]="galleryOptions"
    [images]="galleryImages"
  ></ngx-gallery>
</div>
```

4. Hang-, illetve videotartalmak megjelenítése esetén az alapértelmezett HTML 5-ös videolejátszó kerül felhasználásra, melyet kiegészítettünk néhány extra funkcióval, például a következő, illetve előző hang- vagy videofájlra történő léptetés funkciójával és a lejátszási lista megjelenítésével. Az alábbi kódrészlet az audiovizuális komponens inicializációját tartalmazza, modul szintjén:

```
src\app\+item-page\media-viewer\media-viewer-video\
media-viewer-video.component.ts
```

```
ngOnInit() {
  this.isCollapsed = false;
  this.filteredMedias = this.medias.filter(
    (media) => media.format === 'audio' || media.format === 'video'
  );
}
```

5. Az előző pontban említett audio-, illetve videofájlokat megjelenítő komponens HTML-kódban történő inicializálását az alábbi kódrészlet tartalmazza:

```
src\app\+item-page\media-viewer\media-viewer-video\
media-viewer-video.component.html
```

```
<video
  #media
  [src]="filteredMedias[currentIndex].bitstream._links.content.href"
  id="singleVideo"
  [poster]="
    filteredMedias[currentIndex].thumbnail ||
    replacements[filteredMedias[currentIndex].format]
  "
  preload="none"
  controls
></video>
<div class="buttons" *ngIf="filteredMedias?.length > 1">
  <button
    class="btn btn-primary previous"
    [disabled]="currentIndex === 0"
    (click)="prevMedia()"
  >
    {{ "media-viewer.previous" | translate }}
  </button>

  <button
    class="btn btn-primary next"
    [disabled]="currentIndex === filteredMedias.length - 1"
    (click)="nextMedia()"
```

```
>
  {{ "media-viewer.next" | translate }}
</button>
<div ngbDropdown class="d-inline-block">
  <button
    class="btn btn-outline-primary playlist"
    id="dropdownBasic1"
    ngbDropdownToggle
  >
    {{ "media-viewer.playlist" | translate }}
  </button>
  <div ngbDropdownMenu aria-labelledby="dropdownBasic1">
    <button
      ngbDropdownItem
      *ngFor="let item of filteredMedias; let i = index"
      class="list-element"
      (click)="selectedMedia(i)"
    >
      {{ item.bitstream.name }}
    </button>
  </div>
</div>
</div>
```

6. A médiamegjelenítővel kapcsolatos fejlesztéseket, aktiválásukhoz, a *DSpace 7 front-end* projektjében (DSpace-Angular projekt) engedélyezni kell. Ehhez a lépéshez az *environment.common.ts* fájlban a *mediaViewer*hez kapcsolódó beállításoknál képhez és audiovizuális tartalmakhoz külön-külön engedélyezhető az általunk készített kiegészítő.

Kódrészlet engedélyezése az *environment.common.ts* fájlból:

```
mediaViewer: {
  image: false | true,
  video: false | true
}
```

Az engedélyezéshez a *true* érték megadása szükséges, inaktíválás esetén használjuk a *false* értéket.

Mivel a *DSpace* fejlesztői fontos hozzáadott értékként tekintenek az általunk fejlesztett médiamegjelenítőre, a soron következő *DSpace*-verziókban a modul már valószínűleg alapértelmezetten is „bekapcsolt” állapotban fog rendelkezésre állni.



## Tamás Kiss

*Monguz/Qulto International*

*Eötvös Loránd University Centre for Digital Humanities*

tkiss@monguz.hu

## Zoltán Kanász-Nagy

*Monguz/Qulto International*

zkanasznagy@monguz.hu

# Record Page Media Player in *DSpace* 7

In recent years, public collections and academic research communities have shown a growing interest in open-source repository solutions that are easy to install, customize, and operate. *DSpace*, a repository software developed by the not-for-profit DuraSpace organization since 2002, and its current version 7 proves to be an adequate solution to these requirements regarding storing and repositing digital objects. However, *DSpace* 7 is currently available only in a beta version and lacks certain features and functionalities, the development of which – due to the urgency of the momentary needs that our team intended to meet – could not wait until their planned implementation by the DuraSpace community. A media viewer and media player were such missing functionalities. Therefore, the authors of this article and their colleagues (web designer Annamária Tápai-Kovács and software developer Dániel Sipos) have undertaken to develop these functionalities and offer the resultant blocks of code to DuraSpace to be implemented in the code base of *DSpace* 7. In this paper we shall discuss the brief history of our collaboration with DuraSpace and share code snippets from our development.

Keywords:

DSpace 7, software development, media viewer, media player, repository, Angular 10



## Community Development of *DSpace* 7

*DSpace* is an open-source repository software that has gained great popularity in recent years, primarily, among public education and public collection institutions. It owes its popularity mainly to its relatively easy usage, configurability and operability. Many projects have been implemented thanks to the *DSpace* repository system, and

drawing on the experience of these projects, several – commercial and freely available – add-ons have now become added value to the core software.

*DSpace's* version 7 continues to be developed on the basis of the ideas and needs of its developer and user community,<sup>1</sup> but it is also a new milestone in the life of the software. While *DSpace 7* inherits useful features of its previous versions, partially diverting from the traditions of previous versions, it receives a completely new, web-based user interface – written in the *Angular 10* framework.<sup>2</sup> In addition, several new features beyond front-end development (e.g. REST API and configurable entities) have been added to the list in which DuraSpace outlined the foundations of the new version and the long-term goals of its development.

Tim Donohue, the chief technology officer of *DSpace's* development, originally planned the first release, as the result of the work that began in 2018, to be completed in the summer of 2020. However, due to unforeseen resource shortages, the developers have changed the date of completion of the current beta version to the second quarter of 2021. Conclusively, users will have to wait a few more months until the release.<sup>3</sup>

However, this does not mean that the software cannot be used in its current state: most of the final features are already available in it, and anyone who wishes to use *DSpace 7* to satisfy their own or their customers' repository-related needs before the first release can try the official demo<sup>4</sup> or install the latest version of the software.

## Our Contribution to the Development of *DSpace 7*

Our team started to follow the life of the new *DSpace* version in the first half of 2020, at the time of the publication of the first alpha version,<sup>5</sup> which was free to be tested by the general public. At first, our developers set out to learn the potentials of the new version as testers, and then they tried to give as much feedback as possible to the software's developers. Later, we started to help with the completion of the first version with smaller improvements, as a result of which Monguz/Qulto International, our wider professional „family” – now a member of the DuraSpace developer community – contributed to the development of *DSpace* by fixing various bugs, such as ones affecting displaying search results and record pages.

<sup>1</sup> At an organizational level, the development of *DSpace* is coordinated by DuraSpace, a not-for-profit community, which finances its operation not from the sale of the software but from the financial contributions of its community members. The most formidable member institutions supporting the community financially as well, are listed on the following website, accessed at: 2021.04.27, <https://duraspace.org/dspace/community/members/>.

<sup>2</sup> A detailed description of the back-end and front-end technologies used in the development available at the following link, accessed at: 2021.04.27, <https://wiki.lyrasis.org/display/DSDOC7x/Installing+DSpace>.

<sup>3</sup> Tim Donohue et al., „On the Road to *DSpace 7*: 2018 Edition,” 2018, [http://repositorio.concytec.gob.pe/bitstream/20.500.12390/46/1/2018\\_Donohue\\_Dspace7.pdf](http://repositorio.concytec.gob.pe/bitstream/20.500.12390/46/1/2018_Donohue_Dspace7.pdf).

<sup>4</sup> The demo version is accessible through the following link, accessed at: 2021.04.27, <https://dspace7-demo.atmire.com/home>.

<sup>5</sup> Members of the DuraSpace community communicate with each other not only at larger-scale *DSpace* Steering Group meetings and developer-level technical status discussions, but also through their community chat channel on Slack, to which anyone is free to sign up.

DuraSpace's development roadmap scheduled displaying and playing image, audio, and video content directly on the record page through a web interface only to a later phase of development (late 2021 at the earliest). Consequently, our team undertook the development of functionalities that would enable viewing and playing record-related media, thus furthering the user-friendly operation of the software. Our developers' new implementations will be part of the system in the next (beta 5) version. Those interested in the chronology of this development, which greatly facilitates the proper use of the software, as well as in the details of its implementation, are advised to consult the following Github pages: <https://github.com/DSpace/dspace-angular/issues/885>, <https://github.com/DSpace/dspace-angular/pull/888>.

## Publikáció: Csabai képek



### Leírás

Csabai képek kiállítása..

### Kulcsszavak

Plakát, kiállítás

### URI

<http://repository.qulto.eu/handle/12234>

### Állományok

rekord00000.jpg (5.26 MB)

### Gyűjtemények

Szövegek, képek...

### Kiadó

Békés Megyei Könyvtár

Teljes tárgy oldala

1. Figure. Video player (record page excerpt)

## Development Methodology with Media Player Code Snippets

The development methodology we used for displaying image, audio and video content on the record page and the resultant code snippets are as follows:

1. Preparing record page media display: analyzing the descriptive records of all of the bitstreams associated with the *DSpace* record, we create an object of the type *mediaViewerItem*, that is, we map the bitstream resource descriptor to *mediaViewerItem*:

```
src/app/+item-page/media-viewer/media-viewer.component.ts

/**
 * This method loads all the Bitstreams and Thumbnails and converts
 * them to media item
 */

ngOnInit(): void {
  this.mediaList$ = new BehaviorSubject([]);
  this.isLoading = true;
  this.loadRemoteData('ORIGINAL').subscribe((bitstreamsRD) => {
    if (bitstreamsRD.payload.page.length === 0) {
      this.isLoading = false;
      this.mediaList$.next([]);
    } else {
      this.loadRemoteData('THUMBNAIL').subscribe((thumbnailsRD) => {
        for (
          let index = 0;
          index < bitstreamsRD.payload.page.length;
          index++
        ) {
          bitstreamsRD.payload.page[index].format
            .pipe(getFirstSucceededRemoteDataPayload())
            .subscribe((format) => {
              const current = this.mediaList$.getValue();
              const mediaItem = this.createMediaViewerItem(
                bitstreamsRD.payload.page[index],
                format,
                thumbnailsRD.payload && thumbnailsRD.payload.page[index]
              );
              this.mediaList$.next([...current, mediaItem]);
            });
        }
        this.isLoading = false;
      });
    }
  });
}
```

2. We use the *Ngx-gallery* module modified by Kolkov to display images.<sup>6</sup> For the gallery to work, one needs to define three different image versions for each image file: a small and a medium-sized thumbnail, and a large image – for the latter, we use the originally uploaded image file. The code block below

<sup>6</sup> Andrey Kolkov, „Ngx-gallery,” 2020, accessed at: 2021.04.27, <https://www.npmjs.com/package/@kolkov/ngx-gallery>.



demonstrates that if a thumbnail was uploaded together with the original image file, it will be used by the media viewer. However, the media viewer will use a thumbnail generated by *DSpace*, if one is available. In the absence of any thumbnail, we provide a general-purpose thumbnail (*replacement\_image.svg*).

```
src\app\+item-page\media-viewer\media-viewer-image\  
media-viewer-image.component.ts
```

```
/**  
 * This method convert an array of MediaViewerItem into  
 * NgxGalleryImage array  
 * @param medias input NgxGalleryImage array  
 */  
convertToGalleryImage(medias: MediaViewerItem[]): NgxGalleryImage[] {  
  const mappadImages = [];  
  for (const image of medias) {  
    if (image.format === 'image') {  
      mappadImages.push({  
        small: image.thumbnail  
          ? image.thumbnail  
          : './assets/images/replacement_image.svg',  
        medium: image.thumbnail  
          ? image.thumbnail  
          : './assets/images/replacement_image.svg',  
        big: image.bitstream._links.content.href,  
      });  
    }  
  }  
  return mappadImages;  
}
```

3. Then we display the resulting image list with the aforementioned Kolkov *Ngx-gallery* module:

```
src\app\+item-page\media-viewer\media-viewer-image\  
media-viewer-image.component.html
```

```
<div [class.change-gallery]="isAuthenticated$ | async">  
  <ngx-gallery  
    class="ngx-gallery"  
    [options]="galleryOptions"  
    [images]="galleryImages"  
  ></ngx-gallery>  
</div>
```

4. When displaying audio or video content, the default HTML5 video player is used, which we supplemented with some extra features, such as a functionality

that skips to the next or previous audio or video file and another one displaying the relevant playlist. The following code snippet contains the initialization of the audiovisual component at the module level:

```
src\app\+item-page\media-viewer\media-viewer-video\  
media-viewer-video.component.ts
```

```
ngOnInit() {  
  this.isCollapsed = false;  
  this.filteredMedias = this.medias.filter(  
    (media) => media.format === 'audio' || media.format === 'video'  
  );  
}
```

5. The HTML initialization of the component displaying audio or video files mentioned in the previous section is contained in the following code snippet:

```
src\app\+item-page\media-viewer\media-viewer-video\  
media-viewer-video.component.html
```

```
<video  
  #media  
  [src]="filteredMedias[currentIndex].bitstream._links.content.href"  
  id="singleVideo"  
  [poster]="  
    filteredMedias[currentIndex].thumbnail ||  
    replacements[filteredMedias[currentIndex].format]  
  "  
  preload="none"  
  controls  
></video>  
<div class="buttons" *ngIf="filteredMedias?.length > 1">  
  <button  
    class="btn btn-primary previous"  
    [disabled]="currentIndex === 0"  
    (click)="prevMedia()"  
  >  
    {{ "media-viewer.previous" | translate }}  
  </button>  
  
  <button  
    class="btn btn-primary next"  
    [disabled]="currentIndex === filteredMedias.length - 1"  
    (click)="nextMedia()"  
  >  
    {{ "media-viewer.next" | translate }}
```

```
</button>
<div ngbDropdown class="d-inline-block">
  <button
    class="btn btn-outline-primary playlist"
    id="dropdownBasic1"
    ngbDropdownToggle
  >
    {{ "media-viewer.playlist" | translate }}
  </button>
  <div ngbDropdownMenu aria-labelledby="dropdownBasic1">
    <button
      ngbDropdownItem
      *ngFor="let item of filteredMedias; let i = index"
      class="list-element"
      (click)="selectedMedia(i)"
    >
      {{ item.bitstream.name }}
    </button>
  </div>
</div>
</div>
```

6. Our developments related to the media viewer need to be enabled in the *DSpace 7* frontend project (i.e. *DSpace-Angular* project) in order to become activated. Our developments can be enabled separately for image and audiovisual contents in the settings of media viewer in the *environment.common.ts* file.

Enabling the code from the *environment.common.ts* file:

```
mediaViewer: {
  image: false | true,
  video: false | true
}
```

In order to enable the media viewer one needs to set the relevant value(s) to *true*. In case one intends to fully or partially deactivate the media viewer, they need to set the relevant value(s) as *false*. As the developers of *DSpace* see serious added value in the media viewer, the module is likely to be available as activated by default in future versions of the software.



<KRITIKA>



## Maróthy Szilvia

BTK Irodalomtudományi Intézet

mthy.szilvi@gmail.com

Olivier Le Deuff. *Digital Humanities: History and Development*. In-formation Systems, Web and Pervasive Computing Series, Intellectual Technologies Set 4. London: Wiley-ISTE, 2018. ISBN 9781786300164, 9781119308195. xvi, 150 oldal. <https://doi.org/10.1002/9781119308195>

Olivier Le Deuff *Digital Humanities: History and Development* című kötete a digitális bölcsészet önértelmező, önnön hagyományát kereső munkái sorába illeszkedik. A szerző már a bevezetőben leszögezi, hogy nem a digitális bölcsészet átfogó történetét, vagy geneziséét kívánja megírni, hanem annak összetettebb történeti beágyazottságára reflektál. A digitális bölcsészet „hosszú történetét” (ix), annak fontos állomásait vizsgálja példákon keresztül, melyek szerencsés módon olykor az angolszász szakirodalmon túlról, francia, olasz, német nyelvterületről érkeznek. A megközelítés a hazai szakirodalomban sem egyedülálló, legutóbb például Kokas Károly, Horváth Iván, illetve Z. Karvalics László mutatta be egy-egy cikkében, az időben meddig nyúlhatnak vissza a ma digitális bölcsészet ernyőfogalma alá került kutatások, metodológiák.<sup>1</sup>

A „pre-digitális bölcsészet” történetének bemutatása nem tét nélküli. Kiemelt célja, hogy bebizonyítsa a diszciplína kialakulása, főbb irányai korántsem előzmény nélküliek, így ezek nem hagyhatók figyelmen kívül. (xii.) Köztudott, hogy a 2004-ben a *Companion...*<sup>2</sup> óta széles körben elterjedté vált *digital humanities*, digitális bölcsészet nem csupán egy diszciplínát jelölő fogalommmá vált, hanem olyan kulcsfogalommmá, divatszóvá, mely köré intézményeket (tanszékeket, központokat, folyóiratokat és konferenciákat) lehet építeni. Ez az építkezés pedig sokak szerint csak akkor lehet sikeres, ha az újdonság, újszerűség kerül a középpontba, mely a múlt elfeledésével jár. Le Deuff kötete ezzel a múltat feledő folyamattal megy szembe – a *Digitális Bölcsészet* folyóirat szerkesztőjeként ezt magam is fontosnak tartom, így nem véletlenül esett erre a kötetre a választás.

A „pre-digitális bölcsészet” megnyilvánulásait Le Deuff a *Respublica litterariától* a kvantitatív jellegű kutatásokon, a kézi és automatikus (dokumentum)feldolgozáson

<sup>1</sup> Kokas Károly, „Digitális bölcsészet 2016: A bölcsészek és az informatikai megközelítés: régen és most,” in Nyerges Judit, Verók Attila, és Zvara Edina, szerk. *MONOKgraphia: Tanulmányok Monok István 60. születésnapjára* (Budapest: Kossuth Kiadó, 2016), 405–412; Horváth Iván, „Digitális bölcsészet a virtuális nemzeti könyvtárban,” *Digitális Bölcsészet* 1. sz. (2018): 121–138, <https://doi.org/10.31400/dh-hun.2018.1.234>; Z. Karvalics László, „Libro de los Epitomes: Kutatói kommentárok egy lappangó kéziratok könyv fellelészének híréhez,” *Könyvtári Figyelő* 65, 2. sz. (2019): 260–271.

<sup>2</sup> Susan Schreibman, Ray Siemens, and John Unsworth, eds., *A Companion to Digital Humanities* (Oxford: Blackwell, 2004), <http://digitalhumanities.org/companion/>.

át a térképek készítéséig mutatja be. Így joggal merülhet fel a kérdés, vajon ezek valóban mind a digitális bölcsészet mint diszciplína előzményei-e. Vagy inkább egyfajta megközelítésmódé, mely minden diszciplínának lehet sajátja, némelyiknek jobban, némelyiknek kevésbé?

A szerző feltehetőleg ez utóbbit kívánja megragadni, tehát a sajátos megközelítésmódra hoz különféle területekről példákat. Könyvének egyik hőse Gottfried Wilhelm Leibniz, aki egyszerre volt filozófus, matematikus és könyvtáros, s aki számára egyaránt volt fontos a tudáshoz való hozzáférés és a tudásrendszerezés. (9.) Azt hiszem, kevés olyan könyv van, ami ilyen melegszívűen és magasztosan közelít a könyvtáros szakmához, a kötetben a tudós könyvtárosok végig fontos szerepet kapnak.

Az első, *Republica litteraria*-val foglalkozó fejezet felvonultatja azokat a jellemzőket, melyek Le Deuff szerint a digitális bölcsészet művelőire is érvényesek. Ilyen a tudóshálózat megléte, a virtuális közösségé, melynek tagjai tudást, forrásdokumentumokat osztanak meg egymással és közös nyelvük van. Leibniz levelét idézi, melyben a wolfenbütteli herceg támogatását kéri a könyvtár bővítéséhez. „Akármilyen gyönyörű is egy könyvtár, nem őrizhetjük anélkül, hogy ne gondoskodnánk bővítéséről.” (10.) Leibniz a tudományban való előrehaladás, naprakészség feltételeként szabja az újabb kiadványok beszerzését. E példával Le Deuff egyúttal a mai kutatók információhozzáférési gondjaira is utal, az Open Access mozgalmat itt és a könyv más helyein is felhossa. De hogy köthető ez az inkább könyvtártörténeti adalék a digitális bölcsészethez? Le Deuff szerint a tudásszervezésnek azért fontos állomása Leibniz levele, mert a könyvtár növelésének célkitűzése valójában a tudás növelésére irányul, tehát a tudás külső forrásokkal való bővítésére. (11.)

Az írás és leírás, dokumentáció tudományával foglalkozik a második fejezet, melynek fő állítása, hogy az írás maga is kódolás és programozás, a digitalizálás pedig nem dematerializálás, hanem rematerializálás (16). Ezt Le Deuff azért hangsúlyozza, mert a számítógépes, kivált internetes médiumok esetében gyakorta elfeledkeznek arról, hogy az e médium közegébe került szöveg éppúgy kiadás, annak minden textológiai sajátosságával, mint a kéziratos vagy nyomtatott szöveg. E helyt röviden a megőrzés problematikája is előkerül, mely egyébként alig jelenik meg a kötetben: Le Deuff felhívja rá a figyelmet, hogy nemcsak kulturálisan évílhet el, válhat értelmezhetetlenné a forrás, hanem technikailag is. Megállapításai nem új keletűek, hiszen az elmúlt három évtized szövegdigitalizálási hullámai több ízben felszínre hozták ezt a kérdéskört, ám annak jó összefoglalását adja. A hazai szakirodalomban legátfogóbban talán Horváth Iván munkásságában találkozhatunk ezzel a problematikával. A fejezet ezután az írástudomány és a bibliológia 20. század elején-közepén megjelent új diszciplínáival foglalkozik, különösen Paul Otlet munkásságával.

A harmadik fejezet az előbbi témák folytatásaként a tudásszervezés formáival, a táblázatok, indexelés kérdéskörével foglalkozik. „Megközelítésünk szerint a digitális bölcsészet az indexszel kezdődött.” (23.) Ha a dokumentumokat kimerítően nem is tudjuk elolvasni, legalább tudjunk benne keresni és kutatni (30) – ez, a digitális bölcsészetre valóban jellemző elgondolás az indexek készítésében gyökerezik. Ezt a gondolatmenetet fűzi tovább a negyedik fejezet, mely az információvisszanyerés módozataival foglalkozik.



Az ötödik fejezet elkanyarodik a hipertext története felé, és Vannevar Bush híres cikkének előzményeit tárja az olvasók elé.<sup>3</sup> Ramelli könyvkerekén túl (48) ilyen Herbert George Wells több, Le Deuff által részletesen ismertetett írása. Wells világelme (*world brain*), vagy világciklopédia utópiáját az Egyetemes Dokumentációs Világkongresszuson (World Congress of Universal Documentation) mutatta be 1937-ben Párizsban. Le Deuff a decentralizáltan és mindenki számára hozzáférhető tudás gondolatát emeli ki. Wells elképzeléseit párhuzamba állítja a Világkongresszushoz sok szálon kapcsolódó Paul Otlet munkásságával és Emanuel Goldberg mikrofilmes készülékével (Microfilm Rapid Selector). Ez utóbbi volt az első olyan készülék, mely félautomatikus indexelést, rendszerezést tett lehetővé. (51–53.)

Míg Leibniz a digitális bölcsészet szellemiségét vetíti előre, s ekként bukkan fel a könyvben újra meg újra, addig Paul Otlet, a „proto-digitális bölcsész” már a diszciplína főbb ösvényeit kitaposóként jelenik meg. Az Otlet által felvetett bibliológia, valamint az információfeldolgozással, -kezeléssel és visszakereséssel kapcsolatos elképzelései Le Deuff számára nemcsak a könyvtártudomány, hanem a digitális bölcsészet gyökerét is jelentik. (Ez elvezet ahhoz a gyakorta könyvtáros és bölcsészeti oldalról is felmerülő, hol testvéries, hol a testvériséget megtagadó vitához, hogy a két tudomány azonos-e, rokon-e egymással. Ezt azonban nem célok vizsgálni.) Paul Otlet *Mondothèque* nevű multimédiás dolgozóasztalát szellemiségében Le Deuff az internetéhez hasonlítja. (56–57.) Otlet olyan munkaállomás kialakításán dolgozott, mely egy helyre gyűjti a világ összes információját (alapvetően a nyomtatott forrásokra gondolva). Az információk közti eligazodást szolgálta például az Egyetemes Tizedes Osztályozás (ETO) is, melynek kialakításában Otletnek fontos szerepe volt. Hasznos összefoglaló táblázat található a kötet 58–59. oldalán, mely egybegyűjti az internet, illetve a hipertext történetének főbb állomásait 1910-től 1989-ig.

A hatodik fejezet a társadalomtudományok és a statisztika területére koncentrál, mely elsősorban a digitális bölcsész számítástudományi készségeit, illetve előképeit hivatott bemutatni. Ezért a digitális írástudó definíciójával indít, mely megfogalmazásában azt a személyt jelöli, aki „szakértelmét számolási és számítási módszerekkel képes kibővíteni” (65). A fejezet Émile Durkheim, Lucien Febvre (és az *Annales*), Emmanuel Le Roy Ladurie, François Furet (és a big data gondolata 1959-ből) munkásságát állítja a középpontba, de megjelenik például az adatvizualizáció, adatábrázolás első meghatározó alakja, Jacques Bertin is (76).

A hetedik fejezet az automatikus szövegfeldolgozás, a korpuszpépítés, a konkordanciajegyzékek, illetve a kapcsolt adatok előzményeit foglalja össze. Szóba kerül Roberto Busa, a digitális bölcsészet mitizált ősforrása is. Itt többek közt azt a Tito Orlandival készült interjút idézi, mely nemrég magyar fordításban is napvilágot látott.<sup>4</sup> Orlandi és mások vitatják Busa elsőbbségét az automatikus szövegfeldolgozás terén, s többek között Jean-Claude Gardin munkásságára irányítják helyette a figyelmet.

A következő, metaadatokkal foglalkozó fejezet egy másik érzékeny kérdéskörre reflektál: programozó, IT-szakember-e a digitális bölcsész? „[A] digitális bölcsészeknek nem kell feltétlenül számítógép specialistákká válniuk, inkább kódolókká és dekó-

<sup>3</sup> Vannevar Bush, „As We May Think,” *The Atlantic Monthly* 176, 1. sz. (1945): 101–108.

<sup>4</sup> Tito Orlandi és Julianne Nyhan, „hic Rhodus, hic salta! Tito Orlandi és Julianne Nyhan beszélgetése,” ford. Maczelka Csaba, *Helikon* 66, 1. sz. (2020): 5–15.

dolókká.” (93.) A továbbiakban ennek szellemében újabb izgalmas könyvtártörténeti adalékok következnek a katalogizálásról, az ETO-ról, a taxonómiákról és folkszonómiákról (vagy folkonómiákról), valamint szóba kerül a *web of data* és a gráfadatbázisok témaköre is.

A kilencedik fejezet a bibliometria és a tudománymetria területén tett kitérő, mely a könyvben hangsúlyos világtudás és tudományos közösségek tematika mentén kerül a képbe. Le Deuff ezúttal is hangsúlyozza a digitális bölcsészeti kutatások tudományköziségét, mint óriási pozitívumot.

A kötet tizedik fejezete a valós és virtuális világtérképek felé kanyarodik el, valamint a világtudás modern megjelenítői, Michel Foucault és Jorge Luis Borges felé. Borges ezúttal egy igen rövid novellájával kapcsolódik ide (Del rigor en la ciencia, „A tudomány rigorózusságáról”), mely egy meg nem nevezett (Le Deuff által tévesen, talán az enciklopédiás novella hatására kínainak titulált) birodalom térképészéről szól, s a tökéletes térkép létrehozásáról. Érdekes módon itt, és nem a társadalomtudományokat átfogó hatodik fejezetben kerül tárgyalásra a közösségi hálózatok kutatása, a szociometria és Lev Manovich Instagram-elemzése (118). A könyv összegzésében Le Deuff szintén a „könyvtáros” és „pre-digitális bölcsész” Leibnizre, valamint Foucault gondolataira támaszkodva érvel a tudományköziség mellett. A tudományágak és -területek szerinti gondolkodás időszaka után a digitális bölcsészet lehetősége és egyben feladata a tudományágak és a tudás közötti kapcsolatok (linkek) újraalkotása. (124.)

Noha gyakorta lehet az az érzésünk, hogy Le Deuff valójában nem a digitális bölcsészt, hanem az információelmélet és -technológia, a számításelmélet iránt fogékony tudóst rajzolja meg, mint ilyen, a könyv nagyon inspiráló olvasmány. Példáiból kitűnik, hogy a szerző számára a digitális bölcsészet a világtudás összegyűjtésére, rendezésére és közreadására irányuló tudós tevékenység kortárs testet öltése, valamint egyensúlykeresés a kvalitatív és a kvantitatív vizsgálódások között. Könyve tulajdonképpen felszólítás mindannyiunknak arra, hogy keressük meg közvetlen tudományos környezetünkben is azokat az elődöket, előzményeket, amelyek hozzásegíthettek a ma digitális bölcsészetnek nevezett, ám igen sokszínű kutatási ág(ak) felismeréséhez, kérdésfelvetéséhez, eszköztárához.

## Király Péter

Göttingen eResearch Alliance

Gesellschaft für wissenschaftliche Datenverarbeitung mbH, Göttingen

peter.kiraly@gwdg.de

## A magyar népzenei adatok története és a (digitális) archiválás lehetőségei

**Bolya Mátyás. *Információelmélet és népzene kutatás. Rendszeralkotás, nyilvántartás, digitális archívum*. Budapest: MTA BTK Zenetudományi Intézet–L'Harmattan Kiadó, 2019. ISBN 9786155167171. 348 oldal.**

Bolya Mátyás népzénész és az MTA BTK Zenetudományi Intézete (MTA BTK ZTI) népzenei archívumának vezetője. A szóban forgó kiadvány doktori disszertációjának<sup>1</sup> könyvváltozata.

A kötet két szorosan összefüggő témát jár körül: milyen tulajdonságai vannak a magyar népzene archivált forrásbázisának (a könyvben ezt főként a szerző munkahelye reprezentálja), és hogyan próbálták ezt a forrásbázist rendszerezni az elmúlt bő száz évben a magyar népzene legjelentősebb kutatói annak érdekében, hogy forráskiadványokban és monográfiákban közreadják a „magyar népzene”. Az idézőjelet azért teszem ide, mert, mint a könyvből kitűnik, a népzene fogalma alatt még Bartók és Kodály sem pontosan ugyanazt értette, az utánuk következő nemzedékek, a külföldi szakirodalom pedig újabb, esetenként a magyar kutatásban kevésbé érvényesülő szempontokat is felvetett. A könyv nyolc oldalon keresztül tárgyalja a definíciós kísérleteket. Ízelítőül néhány olyan szempont, melyekkel egyes kutatók lehatárolták a maguk kutatási területét: muzikológiai megközelítéssel koncentrálnak az esztétikai szempontok alapján meghatározott tiszta forrásra (kihagyva például a műzenei hatásokat), vagy antropológiai megközelítéssel azt tárjuk fel, hogy milyen zene volt használatban a kijelölt társadalmi rétegben; csak a hagyományos gyűjtéssel felvett adatokkal dolgozzunk, vagy vonjunk be kevésbé pontosan adatolt történeti forrásokat (melyek például kéziratos énekeskönyvekben maradtak fenn); csak az énekelt dalokat, vagy a hangszeres zenét is elemezzük. Bolya saját népzenefelfogásának meghatározását három és fél oldalon fejt ki (31–34), amely benyomásom szerint a népzene összetett társadalmi-kulturális jelenségnek tekinti, melynek a tulajdonságai helyenként illékonyak, és amelyre jellemző a folytonos változás, ezért „a népzene a gondolati modellben meghatározott, idealizált környezetben ritkán lelhető fel”. (33.)

<sup>1</sup> Bolya Mátyás, *Információelmélet és népzene kutatás: A magyar népzenei rendszerezés és nyilvántartás vizsgálata adatbázis szemlélettel. Doktori értekezés*. (Budapest: Liszt Ferenc Zeneművészeti Egyetem, 2018), VIII, 310 oldal. Elérhető a Zeneakadémia dokumentumtárában: [https://apps.lfze.hu/netfolder/PublicNet/Doktori%20dolgozatok/bolya\\_matyas/disszertacio.pdf](https://apps.lfze.hu/netfolder/PublicNet/Doktori%20dolgozatok/bolya_matyas/disszertacio.pdf).

Ez a fajta szemlélet tökéletesen megfelel annak az archivátori megközelítésnek, ami abban segít, hogy az elődök eltérő szemléletmódjából fakadó gyűjteményi szabálytalanságokat megfelelően kezelni tudjuk.

A szerző ezután a népzene gyűjtés kezdeteit vázolja, majd jó száz oldalon áttekinti a gyűjtött anyag rendezésének történetét. Számomra érdekes módon az derült ki, hogy Bartók, Kodály, Járdányi Pál, Vargyas Lajos és mások mind-mind némileg máshol húzták meg a népzene határait, másrészt ennek megfelelően más és más szempontokat vontak be az elemzésükbe.

Az MTA BTK ZTI népzenei gyűjteményi egységeit (mint a magyar népzenei kutatások legfontosabb forrásbázisát) a szerző egy kétdimenziós ábrával szemlélteti (103–104), melyen egyfelől megkülönbözteti a népzene három nagy csoportját – hangszeres; vokális, nem strófikus (siratóénekek, mondókák, gyerekjátékok); vokális strófikus –, illetve a gyűjtött anyag részeit (hangfelvételekkel is rendelkező Bartók- és Kodály-rendek; későbbi hangfelvételekkel kiegészülő helyszíni dallamlejegyzések; hangfelvétel nélküli lejegyzések és feldolgozatlan anyag). Az anyag, vagyis a dallamok száma folyamatosan nőtt. Bartók és Kodály közös, 1913-as programalkotó tervezete 5500 dallammal számolt, a Bartók-rend (amit 1940-ig épített) 13 500 tételt, a Kodály-rend (1958) 30 ezret, a jelenlegi gyűjtemény pedig hozzávetőlegesen 250 ezret tartalmaz.

Fontos kitérni egy terminológiai problémára. A népzene kutatás „rend”-nek nevezi a rendszeralkotási kísérletek eredményét, amelyek egyszerre elméleti konstrukciók és tényleges gyűjteményrészek. A könyv (követve a szakirodalmat) a rend, rendszer, illetve rendszerezés, rendezés szavakat szinonimaként használja. Közgyűjteményi háttérrel rendelkező recenzensként a könyv (és feltehetően a népzene kutatás) terminológiája néhány más ponton is szokatlan. Ilyen például a szemléletes „forráslánc” kifejezés, ami párhuzamba állítható a filológia stemmájában szereplő egyes leszármazási ágakkal, de annál némileg szűkebb fogalom. A forráslánc esetében a másolás lényegében a gyűjteményen belül jön létre a feldolgozás részként. Első eleme (vagyis az adat fő forrása) szerencsés esetben a helyszíni hangfelvétel, amiről egyfelől dallam- és szöveglejegyzés, másfelől archivált hangfelvétel készül. A forráslánc további tagjai a további másolatok és a publikációk. Némely esetben egyéb forrás hiányában (vagy netán a forrás megsemmisülése okán) 19. századi kiadványok (az első népzene gyűjtők publikációi), illetve kéziratos énekeskönyvek jelentik a forráslánc első elemét, vagyis az elsődleges forrást. Egy másik ilyen terminológiai sajátosság, hogy a Népzenei Gyűjtemény „leltári gyűjtemény” néven a hangfelvétel nélküli helyszíni lejegyzéseket illeti. Bolya dokumentálja a forráscsoportokat, illetve azok történeti változásait. Sajnos a számokat illetően sok a bizonytalanság, ezeket Bolya igyekszik transzparenssé tenni. Alaposságára jellemző a következő megjegyzés: „... de figyelemre méltó, hogy a Zenetudományi Intézet kézikönyvtárában található példányban ezt az adatot valaki ceruzával ~29 000-re javította...” (106, 103. jegyzet).

A következő fejezet (*A népzenei rendezés gyakorlata*) a népzenei rendszeralkotás történetét, és az egyes rendszerek módszertanát vizsgálja. A könyv alapján az a benyomásom, hogy egyik rendszeralkotó sem volt képes pontosan felmérni a saját (vagy csapatuk) teljesítőképességét, ezért szükségszerűen ezek mind csonka vállalkozások. Bolya pontosan adatolja ezeket a vállalkozásokat, felmérve egyrészt eredményüket, az archívumban őrzött anyagokat, másrészt cikkeik, levelezésük, sőt nem egy esetben

egykori munkatársakkal folytatott interjúk alapján feltárva a munkafolyamatot, a döntések elvi-gyakorlati háttérét. A rendszerek elsősorban valamilyen nyomtatott kiadvány előtanulmányaiként szolgáltak, de némelyikük a pusztán elvi szintű megközelítésmód mellett a fizikai gyűjtemény tényleges rendszerezését is magával vonta. Vannak kutatók, akik lineáris rendezést tűztek ki célul, mások többdimenziós, nem bináris, hanem egymást átfedő „felhőkben” vagy „dallambokrokban” gondolkodtak. A rendszerek a dallamok tulajdonságaiból indultak ki, de a rendszerezők nem ugyanazokat a tulajdonságokat használták. A főbb tulajdonságok a ritmus (pl. ritmika, pódia, metrika, tempo), dallam (pl. hangkészlet, hangterjedelem, hangnem), szöveg (rendezettség, szótagszám), forma (pl. motívum, periódus, refrén), illetve egy magasabb rendű tulajdonság, a stílus. Az egymásra következő rendszerezések sokszor finomítják vagy újragondolják elődeik szempontjait, de esetenként akár teljesen új rendszert is alkothatnak. A rendszeralkotás hosszú folyamat, és néha az alkotó elképzelései menet közben változhatnak. Nevezetes példája ennek, hogy Bartók csak Kodály hatására tulajdonított a ritmikának a saját eredeti elképzeléseinél jelentősebb szerepet a rendszerezésben, aki viszont – egy Bolya által idézett levélrészlet tanúsága szerint – néhány évvel később már kritizálta ezt a szempontot. (A szerző nem tér ki rá, de számomra a kötet egyik legváratlanabb információja az, hogy ez a levélrészlet angol nyelvű. Miért írt vajon Kodály Bartóknak angolul? Az idézet forrásául szolgáló Eörsze László-tanulmány és más források alapján bontakozik ki egy tipikusan 20. századi történet. Bartók és felesége 1940-ben kimennek az Egyesült Államokba, de Bartók első házasságából származó fia, Péter, tanulmányai miatt Magyarországon marad. 1941. december 11-én Magyarország hadat üzen az USA-nak, Bartók Péter december 22-én megy apja után. Kodály december 20-án írja a levelét, hogy azt Péter magával vihesse, és azért angolul, hogy ha esetleg az amerikai hatóságok ellenőriznék, ne kobozzák el gyanús dokumentumként. Péter végül kalandos út után csak áprilisban találkozik apjával, és ekkor adja oda a levelet. Az óvatosság nem volt megalapozatlan. Bartóknak egy nagyjából ugyanabban az időszakban, 1941 vége felé kelt levele fennakadt az angol cenzúrán, azt a hatóságok visszaküldték a feladónak, és Kodály majd csak 1950-ben kapta meg a Bartók-hagyatékot kezelő ügyvédtől.)

A megszületett rendszerek sok mindenben különböznek egymástól. Bartóké igyekezett felölelni a teljes (akkor ismert) anyagot, de használata nehézkes, mivel „az eligazodáshoz tökéletesen ismerni kell Bartók eredeti elképzeléseit” (128). Kodály rendszere jóval egyszerűbb, és ez is a teljes (akkor már 30 ezres) forrásanyaggal dolgozott. Járdányi Pál a Kodály-rendet fejlesztette tovább, de ő inkább tipikus példákat ragadott ki és nem törekedett minden dallam besorolására. A könyv ezen a ponton sajnos nem adja meg az egyes rendezések alapján készült adatbázisokat, ugyanakkor utal számítógépes feldolgozásukra. Általában is jellemző, hogy néhány kivételtől eltekintve nincsenek URL-ek a főszövegben és a lábjegyzetben, még akkor sem, ha kimondottan utal ezek meglétére (pl. „Ezt a formátumot használja a típusrend legújabb internetes publikációja is [*Magyar népdaltípusok példatára*].” [140, 128. jegyzet]). A linkek ehelyett a bibliográfiában találhatók, ugyanakkor a főszöveg és az adatbázis kapcsolata néha még a fejezetek közti lapozgatással és a bibliográfia tanulmányozásával sem világos. Bereczky Jánosnak az új stílusú dallamokat illető rendszerével kapcsolatban rekordazonosítókról, fájlnevekről és fájltranszferről is szó van, de sem a bibliográfia,

sem a névmutató alapján nem találni nyomát az adatbázis elérésének. A rendszerek ismertetése után infografikák szemléltetik az egyes rendszerek által lefedett területeket a gyűjteményrészek és a népzenei kategóriák szerint. Megjegyzendő, hogy ezekről az ábrákról hiányzik az adatmennyiségre vonatkozó információ, így nem világos, hogy a színek és geometriai formák által lefedett területek aránya pusztán jelzésértékű, vagy tényleg megközelítően pontos arányokat jelöl-e (ami a recenzens elvárása az ehhez hasonló infografikákkal szemben). A fejezet végén viszont egy remekül használható, a rendszerek genealógiáját megvilágító infografika áll.

A *számítógép és népzene kutatás* című fejezet a számítógépes alkalmazások történetét vázolja fel. A szerző, ha lehet, itt még az előzőeknél is alaposabb forráskritikával jár el, abból következően, hogy nagyon kevés tényszerű információ maradt fenn. A csekély számú megjelent tanulmány és információmorzsa sok esetben inkább vágyálmokat, terveket közöl, mint tényleges eredményeket. Egy jellemző részlet: „[...] Halmos István adatlapján azt olvashatjuk, hogy tagja a ZTI számítógépes népzene kutatással foglalkozó munkacsoportjának. Nem tudjuk, hogy hivatalosan létezett-e ilyen munkacsoport az intézetben.” (165.) Bolya a források körét tágítandó interjúkat is készített néhány egykori szereplővel, s ezek segítenek megvilágítani az intézménytörténet eme elhanyagolt fejezetét. Annyi megállapítható, hogy feltehetően más intézményekhez hasonlóan az informatika anyagi, tudományos és menedzselési okok miatt is másodrendű szerepet kapott, az informatikusokat (nevezzük most így mindazon szakmák művelőit, akik számítástechnikával foglalkoznak) rövid ideig, általában csak pályájuk kezdeti szakaszában sikerült megtartaniuk. Emiatt a fejlesztések nem épültek egymásra, sokszor alakult ki stagnáló fázis, illetve az új lehetőségeket alkalmasint túlintertézték, és többet vártak tőlük, mint amennyi reálisan elvárható lett volna. A dokumentáció hiányára és az egyik ilyen fellángolásra kiváló példa Víg Rudolf lyukkártyás gyűjteménye, ami a disszertáció írása közben bukkant fel egy egykori résztvevő, Valis Éva adományának köszönhetően. Megjegyzendő, hogy Víg Rudolf fia, Víg Mihály is folytatott cigánynépzene-gyűjtést,<sup>2</sup> talán érdemes lenne őt is meginterjúvolni, hátha további adatokkal szolgálhat a lyukkártyákkal kapcsolatban.

Az *adatbázis-szervezés akadálypályáján* című alfejezet általános szintre emeli a tapasztalatokat, és a néprajzi adatbázisok története alapján néhány fontos következtetést von le az elakadt fejlesztések okairól. Ezt a részt bátran lehet ajánlani a közgyűjteményi digitalizációs projektek döntéshozóinak, Bolya tanácsai segíthetnek elkerülni néhány szóba jöhető csapdát. A fejezet vége a ZTI online publikációs stratégiájával foglalkozik, benne a következő, számomra meglepő mondattal: „Már nem csak néhány szakértő kíváncsi egymás véleményére, komoly igényeket támasztanak például a felsőoktatásban tanuló diákok, a múzeumpedagógusok, az interdiszciplináris beállítottságú gondolkodók stb.” (103–104.) Sajnos, az állítást nem támasztja alá semmilyen adattal, így eldönthetetlen, hogy a szerző (és a recenzens) kívánságáról vagy tényleges folyamatról van szó. Ezzel együtt a vázolt publikációs stratégia akár kutatói-

<sup>2</sup> Marton László Távolodó, „»Ami közelebb hoz« (Víg Mihály),” *Magyar Narancs*, 1998. szeptember 10., [https://magyarnarancs.hu/film2/ami\\_kozelebb\\_hoz\\_vig\\_mihaly-57640](https://magyarnarancs.hu/film2/ami_kozelebb_hoz_vig_mihaly-57640). A gyűjtés menetéről anekdotikusan beszámol a résztvevő, Dénes József. (A *Szökésben* eredetileg, de már a szerző halála után jelent meg a <http://szokesben.blogspot.com/2012/> URL-en, majd rövidítve könyv alakban is [Budapest: Jaffa Kiadó, 2016]).

archivátori életműprogramnak is beillik, teljesülése nagyon fontos hozzájárulás lenne nemcsak a népzene-történeti infrastruktúrához, de az egész online kultúrához. Egy megjegyzést azonban tennem kell. A tervezett platform egyik fejlesztési pontja így hangzik: „DOI-azonosítási lehetőség az adathierarchia bármely pontján.” (190.) Több fajta DOI szolgáltatás létezik (pl. Crossref, DataCite), de közös tulajdonságuk, hogy egyik sem kimondottan olcsó. Viszont a DOI mellett léteznek más elosztott és nem elosztott perzisztens azonosító rendszerek, melyek olcsóbbak, így működési modelljük fenntarthatóbb.<sup>3</sup> Javasolnám egy hibrid modell alkalmazását: DOI-t olyan adatpontokhoz, ahol fontos a metaadatok központi adatbázisban való megjelenése, és egyéb azonosítót ott, ahol elsősorban az adatpont helyének meghatározása a lényeges.

Az utolsó fejezet (*A nem zenei információk*) azokat a főbb metaadattípusokat tárgyalja, melyek nélkül nem lehet kontextusba helyezni a dallamokat, s melyek nélkül elképzelhetetlen lenne számos kutatási irány, például a dialektusok térbeli elterjedésének, az interetnikus hatásoknak, a tánc és zene kapcsolatának, társadalmi szerepének vizsgálata. Bolya ezen belül két nagy csoportot vizsgál: a helyet, időt és személyeket leíró, azonosító adatokat, illetve az úgynevezett járulékos adatokat, ide sorolva a helyszínen gyűjtött további információkat és az archívumi nyilvántartás azonosítóit. (Mellékesen említendő, hogy a fogalmazása kissé pontatlan, hiszen az adatokhoz nyilvánvalóan lehet további, elsősorban az adat tartalmát jellemző, de nem a helyszínen gyűjtött, hanem a kutatás későbbi pontjain hozzáadott járulékos adattípusokat rendelni, sőt néhány ilyen ő maga is felsorol a 204. jegyzetben, mint „egységes angol nyelvű tematikus fogalomtár”, illetve néprajzi fogalmak polihierarchikus thesaurusát.) Ahogy más fejezetek esetében, úgy itt is feltár olyan, nemcsak a saját archívumára vonatkozó tanulságokat, amelyek egyéb közgyűjtemények praxisában is alkalmazhatók, különösen *Az adatkezelés hibái* részben. Néhány ilyen példa: „az archívumot fejlesztő szakember [...] kénytelen a terület szaktekintélyeit az elmúlt évtizedek archiválási gyakorlatának hibáival szembesíteni” (196); „a számítógép csupán egy eszköz, az érték a jól dokumentált adat” (uo.); „[a]z archívumi személyzet leépítésével és átlagos szak-képzettségének csökkenésével párhuzamosan csökken a dokumentálásra fordított idő is. A tapasztalat szerint a dokumentáció helyét előbb-utóbb átveszi a mindennapi rutin, az emberi emlékezet és a szokásjog.” (198.) Legalább ilyen fontosnak tartom a számvetést az archívumi és kutatói szemlélet különbözőségével, valamint az ebből fakadó eltérő prioritásokkal, például ami az anyag referálására vagy szervezésére vonatkozik. Néha megesik más területeken is, hogy a kutatásokkal kiérdemelt tekintély túlzott tisztelete akadályozza az archívumi gyakorlat korszerűsítését. A helymeghatározást illető részeknél a szerző elsősorban két alapproblémára koncentrálna: 1) mit tekintünk történeti-földrajzi referenciapontnak (a korábbi, elsősorban az 1913-as helynévtárra való hivatkozási móddal szemben a térképi geokoordináták, illetve az ezt kiegészítő adatbázisba szervezett és a térképi ábrázolással együttműködő, időpontokhoz kötött névváltozatok használata mellett érvel); 2) hogyan ábrázoljuk a dialektusterületeket (van-e meghúzható határuk, hogy kezeljük a vegyes, vagy tisztázatlan területeket). Ez utóbbira négy online térkép módszerét elemzi, de sajnos a máshol is alkalma-

<sup>3</sup> A közgyűjteményekben használható perzisztens azonosítók kurrens összefoglalását lásd Lukas Coster, „Persistent Identifiers for Heritage Objects,” *Code4Lib Journal* Issue 47, (2020), <https://journal.code4lib.org/articles/14978>.

zott gyakorlat, miszerint internetes forrásokra csak a bibliográfia hivatkozik, kissé kényelmetlenné teszi ezen rész használatát, különösen Bolya saját maga készítette térképe esetében, ahol nem adott meg egy rövid URL-t, hanem egy gyakorlatilag begépelhetetlen hosszúságú Google Maps-linket (a térkép címe „A magyar népzenei dialektusterületek kísérleti térképe” nem ad találatot a Google keresője, az URL pedig még a disszertációból sem másolható ki). Ugyanebben a fejezetben bukkan fel egy érdekes és a *Digitális Bölcsészet* folyóirat számára különösen figyelemreméltónak ígérkező hivatkozás Borsos Balázs kutatására, aki a kilenckötetes *Magyar néprajzi atlasz* adatait szervezte adatbázisba, majd ezeken végzett klaszteranalízist (az egyik első, nagyobb lélegzetű magyarországi digitális bölcsészeti munka – bár készítői természetesen nem is nevezhették ennek – Bácskai Vera és Nagy Lajos 1984-es műve, a *Piackörzetek, piacközpontok és városok Magyarországon 1828-ban*,<sup>4</sup> mely szintén a klaszter- és faktoranalízis segítségével végzett térszerkezeti elemzést). Bolya hivatkozik ugyan Borsos kétkötetes művére, de nem tudjuk meg, hogy az adatbázis elérhető-e valahol. Bolya érdekes módon sem itt, sem a nevekről szóló szakaszban nem tér ki a Nemzeti Névtérre,<sup>5</sup> ami számos itt felvetett problémára (pl. a települések név- és státuszváltozásának kezelése, adatmegosztás, a nevek és névhordozók elválasztása) megoldást nyújt. Bár a szolgáltatás béta verziója csak a könyv előszavának kelte előtt néhány hónappal indult, Ungváry Rudolf és Szakadát István publikációi és előadásai<sup>6</sup> jó néhány évvel ezelőtt ismertették a tervezett szolgáltatás jellemzőit.

A fejezet az információhordozót illető gondolatokkal zárul. A szerző mindenképp előtt igyekszik tisztázni az adat jelentését, ami a korábbi kutatásokban és archiválási gyakorlatban természetesen összekeveredett az információhordozóval és néha a típussal is. Bolya adatnak a bekövetkezett zenei folklóreseményt tekinti, és ennek alapján különíti el az adat egységét az azt hordozó anyagtól. Bár ez az olvasónak 2021-ben triviálisnak tűnhet, ha arra gondolunk, hogy a metaadat-kezelés szempontjából sokkal nagyobb lehetőségekkel bíró könyvtári világban sem következett be az évtizedek óta várt műközpontú, FRBR-alapú katalogizálási gyakorlat, akkor nem csodálhatjuk, hogy Bolya oldalakon keresztül érvel megközelítése mellett rámutatva korábbi meghatározások pontatlanságaira. Apropos pontatlanság – a 234. jegyzetben ezt olvashatjuk: „1950 és 2000 között gyakorlatilag megtízszereződött a gyűjtött anyag”, ezzel szemben a 105. oldalon szereplő, az adatmennyiség változásait ismertető táblázatban 1950-re vonatkozóan 29–35 ezret (lásd az idézett minuciózus 103. lábjegyzetet), 2014-re kb. 250 ezret, a 106. oldalon található grafikonról 2000-re 200 ezer alatti (180–190 ezer) számot olvashatunk. Akármerről nézzük, a növekmény mindössze hat és félszeres.

<sup>4</sup> Bácskai Vera és Nagy Lajos, *Piackörzetek, piacközpontok és városok Magyarországon 1828-ban* (Budapest: Akadémiai Kiadó, 1984).

<sup>5</sup> Nemzeti Névtér, hozzáférés: 2021.01.24, <http://abcd.hu/>.

<sup>6</sup> Ungváry Rudolf, „A névtér, mint kulturális szükséglet: A nemzet aggregátor kitüntetett feladata,” *Tudományos és Műszaki Tájékoztatás* 59, 8. sz. (2012): 320–326, [http://epa.oszk.hu/03000/03071/00057/pdf/EPA03071\\_tmt\\_2012\\_08\\_320-326.pdf](http://epa.oszk.hu/03000/03071/00057/pdf/EPA03071_tmt_2012_08_320-326.pdf); Szakadát István „Magyar Nemzeti Névterek” című előadása a *Networkshop 2016* konferencián: 2016. április 1. <http://kifu.videotorium.hu/hu/recordings/13009>, újabban: Szakadát István, „A névterekről általában,” *Tudományos és Műszaki Tájékoztatás* 67, 1. sz. (2020): 1–11, [http://epa.oszk.hu/03000/03071/00136/pdf/EPA03071\\_tmt\\_2020\\_01\\_001-011.pdf](http://epa.oszk.hu/03000/03071/00136/pdf/EPA03071_tmt_2020_01_001-011.pdf).



Bolya tipológiát épít az információhordozók és tartalomtípusok szerint. A gyűjteményi egységek fizikai tárolását (pl. könyvtár, kéziratár, hangtár) is adatbázisnak, nevezetesen analóg adatbázisnak tekinti – ami számomra meghökkentő, ugyanis nem ismerem olyan adatbázis-definíciót, amely nem foglalja magába a digitális kezelés aspektusát. Nem látom át, hogy a meghatározás ilyen kitágítása milyen elméleti vagy gyakorlati előnyökkel járna, és a könyv sem szolgál erre magyarázattal. Ha mindenképpen kategóriát kell keresnünk, akkor a népzenei archívum anyagára szerencsésebbnek tartanám az adattár megnevezést, amit egyaránt használnak digitális és analóg gyűjteményekre (pl. a *Magyar Nemzeti Múzeum Régészeti Adattára*<sup>7</sup>). Digitális adatbázisok tekintetében megkülönböztet statikust, dinamikus és integráltat (ez utóbbira a „Harvested web content”-et hozza fel példaként, amit érdemes lett volna bővebben kifejteni).

A fejezet az azonosítóra vonatkozó gyakorlati útmutatóval zárul. Bolya értelmezésében az azonosítónak szemantikai jelentést kell hordozniuk. Ennek megvan a maga gyakorlati haszna (erre hoz is példákat), ugyanakkor némileg ellentmond annak a könyv több pontján hangoztatott elvnek, hogy a fájlokat a metaadatokkal együtt kell tárolni – ugyanis ha ez a feltétel megvalósul, akkor a szemantikai rétegnek felesleges az azonosítóban is megjelennie. Bár Bolya nem reflektál rá, de látnunk kell, hogy ez egy régi vita, és például az XML-alapú metaadatsémák általában beszélő neveket használnak (pl. dc:title vagy ead:agencyname, ahol már a névből ki lehet következtetni a jelentést), mások, például a Wikidata, egyszerű sorszámokat (pl. Q235858 a „népzene” azonosítója, ami semmiféle szemantikus viszonyban sincs a sorszámát tekintve előtte álló Q235857-tel, ami Portugália Szent Erzsébet portugál királynőt, II. András dédunokáját jelöli). A sors fintora, hogy a példaként szerepeltetett azonosító (MTA\_BTK\_ZTI\_Mg\_00345) előtagja a megjelenés óta valószínűleg elavult, és az újabb direktíváknak megfelelően le kellene cserélni ELKH-ra.

Egy tévedésre hadd hívjam fel a figyelmet. A 256. lábjegyzetben ez olvasható: „Egyre gyakoribb azonban e logikus sorrend megsértése, talán a kiemelés céljával: [www.btk.mta.hu](http://www.btk.mta.hu).” A domainnevek jelenlegi rendszerét még a világháló főbb szabványainak kifejlesztője, Tim Berners-Lee határozta meg, attól nem tanácsos (nagy vonalakban: el sem lehet) eltérni, tehát ez nem a magyar tudósok felelőssége. A dolog ironiája, hogy Berners-Lee azóta megbánta a döntését, és jelenleg jobbnak látná a nagyobbtól a kisebb felé haladó rendezést.<sup>8</sup>

A könyvet a további kutatási tervek felsorolása, forrásjegyzék, terjedelmes függelék, képek és táblázatok jegyzéke, valamint név- és tárgymutató zárja.

Összességében nagyon fontos és mindenképpen hiánypótló könyv az *Információelmélet és népzene kutatás*. Leginkább egy ritkán művelt dokumentumtípusra, a levéltári ismertetőre<sup>9</sup> emlékeztet, mely a levéltár leendő felhasználóját (a kutatókat, de a levéltárosokat is) bevezeti egy adott levéltár szövevényes történetébe, és is-

<sup>7</sup> *Magyar Nemzeti Múzeum Régészeti Adattára*, hozzáférés: 2021.01.24, <https://mnm.hu/hu/gyujtemenyek/central-database/regeszeti-adattar>.

<sup>8</sup> Például „Interview with Tim Berners-Lee,” *Impact Lab*, <https://www.impactlab.com/2006/03/25/interview-with-tim-berners-lee/>.

<sup>9</sup> Például *Budapest Főváros Levéltára*, szerk. Kenyeres István és Sipos András, Levéltárismertető 2 (Budapest: Budapest Főváros Levéltár, 1997).

merteti az egyes gyűjteményrészek jellemzőit, illetve kutatási segédleteit. A könyv azonban túlmegy ezen a kereten, és felvázolja az archívum digitális fordulatának jelenleg látható kritériumait, problémáit, megoldási javaslatait. Nagyon hasznosnak tartanám, ha e könyv kedvet csinálna más közgyűjtemények munkatársainak ahhoz, hogy hasonló, a korabeli változó szempontokat és a digitalizálás során figyelembe veendő tényezőket egyaránt feltáró metaadat-történeti kutatásokat végezzenek. Főleg azokban az intézményekben járna ez nagy haszonnal, ahol a gyűjteménytörténet hasonlóan jelentős mértékben befolyásolja a kutathatóságot és a digitalizálást, illetve amelyeket a ZTI archívumához hasonlóan eddig elsősorban csak a szűkebb szakterület kutatói ismertek és aknáztak ki.

Ki kell emelni a kötet gondozását, különösen a példás tipográfiát, ezen belül is főként a margináliákat és a *Forrásjegyzék* betűmutatóit; régen olvastam ilyen szépen szedett könyvet (a tipográfia Szalay Éva munkája). A kiadó nyelvi lektort is alkalmazott Lehotka Gábor személyében. A kiadással kapcsolatban két dolgot látok problémásnak. A terjedelmes függelék miatt a *Forrásjegyzék* nem esik kézre, szerencsésebb lett volna a könyv végére, közvetlenül a mutatók elé tenni, illetve a (Demeter Gitta által készített) infografikák színesben jobban érvényesültek volna.

A kötet hátránya – noha ismereteim hiányában nem tudom megítélni, hogy a szöveget vagy a népzenei archívum gyakorlatát kell-e kritizálnom –, hogy nagyon kevésbé reflektál a saját, szűken vett világán kívül eső fejleményekre. Többször is felbukkan az 1964-es *International Folk Music Council* konferencia, de annak ellenére, hogy a szervezet azóta is aktív (2019-ben volt a 45. konferenciája az 1983-ban *International Council for Traditional Music*-ra átnevezett szervezetnek), úgy tűnik, hogy a későbbi fejlemények nem gyakoroltak jelentős hatást az archívumra. A korai, még Bartók által követett példák kivételével szinte egyáltalán nem szerepel a könyvben az, hogy más európai országokban milyen alapon dolgozták fel a népzenei, vagy hogy egyáltalán feldolgozták-e. A kötetben nem látszik sok nyoma annak, hogy más zenei archívumok (pl. zenei könyvtárak, kottatárak, zenekiadók) hogyan működnek, milyen rendszereket építettek, milyen adatbázis-szerkezetet és milyen szabványokat használnak, holott részben legalábbis hasonló problémákat kell megoldaniuk. Néhány könyvtártudományi írás található ugyan a forrásjegyzékben, de érzésem szerint több fontos párhuzam hiányzik, például az egységes besorolási adatokra vonatkozó bő-séges, magyar nyelven is elérhető irodalom és szabványcsalád, amit az adatbázis-tervezés során lehetne hasznosítani. A személynevek problémájára a levéltári világban is találni megoldást (pl. *Encoded Archival Context*).<sup>10</sup> Az idő kódolására minden közgyűjteménytípusnak megvan a maga – esetenként mind a formalizáltság jellegében, mind szemantikájában eltérő – eljárása, most csak egy W3C-ajánlásra hívom fel a figyelmet: *Time Ontology in OWL*,<sup>11</sup> amely különösen a bizonytalan dátumok kódolásához adhat hasznosítható ötleteket. A szöveg és a forma jellemzőinek feltárásához érdemes lenne az irodalomtörténeti megközelítéseket is figyelembe venni, például a Horváth Iván és munkatársai által készített *RPHA (Répertoire de la poésie hongroise*

<sup>10</sup> Encoded Archival Context for Corporate Bodies, Persons, and Families (EAC-CPF), hozzáférés: 2021.01.24, <https://eac.staatsbibliothek-berlin.de/>.

<sup>11</sup> Simon Cox and Chris Little, eds., *Time Ontology in OWL*. W3C Candidate Recommendation 26 March 2020, hozzáférés: 2021.01.24, <https://www.w3.org/TR/owl-time/>.

*ancienne*) adatbázist.<sup>12</sup> A könyv címében szereplő információelmélet szót is túlzónak tartom, mivel noha kétségtelenül a népzenei adatbázisok szervezése során felbukkanó problémák megoldásának elméleti megalapozását adja, a Shannon, Weaver és követőik nevével fémjelzett információelmélet nevű tudományterület körébe tartozó kutatási témák nem képezik a kötet tárgyát. A könyv többször ejt szót algoritmusokról is, de egyet sem ismertet konkrétan. Őszintén remélem, hogy a hiányosságok felsorolása nem veszi el a szerző kedvét, inkább tovább ösztönzi őt ambiciózus programjának megvalósításában.

<sup>12</sup> *Répertoire de la poésie hongroise ancienne*, hozzáférés: 2021.01.24, <https://f-book.com/rpha/>. Az RPHA adatbázisszerkezet részletes ismertetését lásd Horváth Iván, *Gépeskönyv* (Budapest: Balassi Kiadó, 2006), 27–79.